

MERCK PATENT GmbH (DE)	
FEATURES	Location/Qualifiers
source	1. 1305
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
CDS	1. >1305
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAC28479.1"
	/db_xref="GI:12710816"
	/db_xref="REMTREMBL:CAC28479"
	/translation="MDPSDQPLNSLDKPLRPRIPMETFRKVGPIIILSLIASI ILVVLIKVLIDKYFLCGPLHFIPIKQICDGLCEDEBHCVKSPPEGPFAV RLSDRLTQVLDSATGNWFSACFDNFTEALAEATACRQMGYSKPTFRVAFVIGPDODL DVVEITNSQELRMNSGSPCLSGSLVSLHCLACGSKLTPRVVGGSEASVSPQVQ SIOYDKHVGSGSLDHPHWLTAACHPRKHTDVFNKVRAGSKLGSFPLAVAKII IEFNWPKNDIALMKLQPLPSTFVRPICLPFDEBLTPATPKILGWGFTKONG GMSDILQASVQVIDTRCNADDAVOGEVTEKMKCAGIPEGGVDTTCQDSSGSLMYQ SDQHHVVGIVSWGYGCGPSTPGVYTKVSAYLNNWVNWKAEL"
ORIGIN	
Query Match	100.0%; Score 1305; DB 6; Length 1305;
Best Local Similarity	100.0%; Pred. No. 7.5e-302;
Matches 1305; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGGATCCTGACAGTATCAACCTCTGAAAGCCTCGATGTCAAAACCCCTGCGCAACCC 60
DB	1 ATGGATCCTGACAGTATCAACCTCTGAAAGCCTCGATGTCAAAACCCCTGCGCAACCC 60
QY	61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATAGCACTACTGAGC 120
DB	61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATAGCACTACTGAGC 120
QY	121 CTGCGAGTATCATATGTGTTGTCTCATCAAGTGTATCTGGATAAATACTACTTC 180
DB	121 CTGCGAGTATCATATGTGTTGTCTCATCAAGTGTATCTGGATAAATACTACTTC 180
QY	181 CTCTCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGACGGAGAGCTGGAC 240
DB	181 CTCTCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGACGGAGAGCTGGAC 240
QY	241 TGTCCCTTGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTG 300
DB	241 TGTCCCTTGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTG 300
QY	301 GCAGTCCGCTCTCAAGGACCGATTCACACTGAGGTCTGGAATCGGCACAGGGAAC 360
DB	301 GCAGTCCGCTCTCAAGGACCGATTCACACTGAGGTCTGGAATCGGCACAGGGAAC 360
QY	361 TGGTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAG 420
DB	361 TGGTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAG 420
QY	421 ATGGCTACAGACAGAAACCACTTTCAGAGCTGTGAGATTTGGCCACAGAGATCTG 480
DB	421 ATGGCTACAGACAGAAACCACTTTCAGAGCTGTGAGATTTGGCCACAGAGATCTG 480
QY	481 GATGTGTTGAATTCACAGAAACAGCAGGAGCTTCGATCGGGAACCTCAAGTGGGCC 540
DB	481 GATGTGTTGAATTCACAGAAACAGCAGGAGCTTCGATCGGGAACCTCAAGTGGGCC 540
QY	541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCATCTGTTCCCTGTGGGAAGAGCCTGAAGAC 600
DB	541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCATCTGTTCCCTGTGGGAAGAGCCTGAAGAC 600
QY	601 CCCGCTGTGTGGTGGGAGGAGCCTCTGTGATTTGGCTTGGCTTGGAGGTCCAGATC 660
DB	601 CCCGCTGTGTGGTGGGAGGAGCCTCTGTGATTTGGCTTGGCTTGGAGGTCCAGATC 660
QY	661 CAGTACGACAAACAGCAGCTGTGGAGGAGGAGCATCTTGGACCCCACTGGCTCTCAG 720
DB	661 CAGTACGACAAACAGCAGCTGTGGAGGAGGAGCATCTTGGACCCCACTGGCTCTCAG 720
QY	721 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGTCA 780
DB	721 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGTCA 780
QY	781 GACAAACTGGGAGCTTCCCATCCCTGGCTGGGCAAGATCATCATATTGAATTCAC 840
DB	781 GACAAACTGGGAGCTTCCCATCCCTGGCTGGGCAAGATCATCATATTGAATTCAC 840
QY	841 CCATGTATCCCAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTACTTTC 900
DB	841 CCATGTATCCCAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTACTTTC 900
QY	901 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 960
DB	901 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 960
QY	961 CCACTCTGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATA 1020
DB	961 CCACTCTGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATA 1020
QY	1021 CTGCTCGAGGCTCAGTCCAGTCAATTGACACACACCGTGCATGACAGATCGGTAC 1080
DB	1021 CTGCTCGAGGCTCAGTCCAGTCAATTGACACACACCGTGCATGACAGATCGGTAC 1080
QY	1081 CAGGGGAGTCAACGAGAGATGTGTGCGAGCATCCCGAAGGGGTGTGACACC 1140
DB	1081 CAGGGGAGTCAACGAGAGATGTGTGCGAGCATCCCGAAGGGGTGTGACACC 1140
QY	1141 TGCCAGGCTGACAGTGTGGGCTTGTACCAATCTGACCACTGTGCTGTGGGC 1200
DB	1141 TGCCAGGCTGACAGTGTGGGCTTGTACCAATCTGACCACTGTGCTGTGGGC 1200
QY	1201 ATCGTGTAGTGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACCAAGGTC 1260
DB	1201 ATCGTGTAGTGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACCAAGGTC 1260
QY	1261 TCAGGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
DB	1261 TCAGGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
RESULT 2	
LOCUS	AR142620 2038 bp DNA linear PAT 08-AUG-2001
DEFINITION	Sequence 18 from patent US 6203979.
ACCESSION	AR142620
VERSION	AR142620.1 GI:15103906
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 2038)
AUTHORS	Bandman, O., Hillman, J. L., Yue, H., Guegler, K. J., Corley, N. C., Tang, Y. I., Rom, and Shah, P.
TITLE	Human protease molecules
JOURNAL	Patent: US 6203979-A 18 20-MAR-2001;
FEATURES	Location/Qualifiers
source	1. 2038
	/organism="unknown"
	/mol_type="unassigned DNA"
ORIGIN	
Query Match	99.9%; Score 1303.4; DB 6; Length 2038;
Best Local Similarity	99.9%; Pred. No. 1.8e-301;
Matches 1304; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 ATGGATCTCTGACAGTATCAACCTCTGAAACGCTCGATGTCAAAACCCCTGCGCAACCC 60
DB	200 ATGGATCTCTGACAGTATCAACCTCTGAAACGCTCGATGTCAAAACCCCTGCGCAACCC 259
QY	61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120

Db 260 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 319
Qy 121 CTGGCAGATCATCATTTGTGTTGTCTCTCATCAAGGTGATTCGGATAAATACTACTTC 180
Db 320 CTGGCAGATCATCATTTGTGTTGTCTCTCATCAAGGTGATTCGGATAAATACTACTTC 379
Qy 181 CTCTGGGGCAGCTCTCCACTTATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 240
Db 380 CTCTGGGGCAGCTCTCCACTTATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC 439
Qy 241 TGTCCCTTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 300
Db 440 TGTCCCTTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 499
Qy 301 GCAGTCCGCTCTTCAAGGACCGATCCACATCTGCAAGGTGCTGGATCGGCGACAGGGAAC 360
Db 500 GCAGTCCGCTCTTCAAGGACCGATCCACATCTGCAAGGTGCTGGATCGGCGACAGGGAAC 559
Qy 361 TGGTCTCTGCTGCTTTCGACAACTTTCAGAACTTTCAGAACTTTCGCTGAGACAGCTGTAGCGAG 420
Db 560 TGGTCTCTGCTGCTTTCGACAACTTTCAGAACTTTCGCTGAGACAGCTGTAGCGAG 619
Qy 421 ATGGCTACAGAGCAAAACCCATTTTCAGAGCTGTGGAGATTGCCCCAGACAGGATCTG 480
Db 620 ATGGCTACAGAGCAAAACCCATTTTCAGAGCTGTGGAGATTGCCCCAGACAGGATCTG 679
Qy 481 GATGTTGTGAAATCACAGAAACAGCCAGAGCTTCCATGCGGAACTCAAGTGGGCC 540
Db 680 GATGTTGTGAAATCACAGAAACAGCCAGAGCTTCCATGCGGAACTCAAGTGGGCC 739
Qy 541 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAGAGAGCTGAAGACC 600
Db 740 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAGAGAGCTGAAGACC 799
Qy 601 CCCGCTGTGGTGGGAGAGAGCCCTCTGTGATTTCTTGGCTTGGAGTGGAGTCAAGTCA 660
Db 800 CCCGCTGTGGTGGGAGAGAGCCCTCTGTGATTTCTTGGCTTGGAGTGGAGTCAAGTCA 859
Qy 661 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGGACCCCTCACTGGGTCTCAAG 720
Db 860 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGGACCCCTCACTGGGTCTCAAG 919
Qy 721 GCAGCCACTCTTCAGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCA 780
Db 920 GCAGCCACTCTTCAGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCA 979
Qy 781 GACAACTGGCAGCTTCCATCCCTGCTGTGGCCAGATCATCATATTGATTCAC 840
Db 980 GACAACTGGCAGCTTCCATCCCTGCTGTGGCCAGATCATCATATTGATTCAC 1039
Qy 841 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTC 900
Db 1040 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTC 1099
Qy 901 TCAGGCAAGTCAAGCCCATCTGTGTGCTTCTTGTGATGAGGAGCTCACTCCAGCCACC 960
Db 1100 TCAGGCAAGTCAAGCCCATCTGTGTGCTTCTTGTGATGAGGAGCTCACTCCAGCCACC 1159
Qy 961 CCACCTCTGGATCATTGGATGGGCTTTACGAAGCAGAAATGAGGAAGATGCTGACATA 1020
Db 1160 CCACCTCTGGATCATTGGATGGGCTTTACGAAGCAGAAATGAGGAAGATGCTGACATA 1219
Qy 1021 CTGCTGAGGCGTCAAGTCCAGTCAATGACAGCACAGGTGCAATGAGAGAGTGCCTAC 1080
Db 1220 CTGCTGAGGCGTCAAGTCCAGTCAATGACAGCACAGGTGCAATGAGAGAGTGCCTAC 1279
Qy 1081 CAGGGGGAAGTCAAGGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGACACC 1140
Db 1280 CAGGGGGAAGTCAAGGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGACACC 1339
Qy 1141 TGCCAGGCTGACAGTGTGGGCCCTGATGTATGATGATGATGATGATGATGATGATGATG 1200
Db 1340 TGCCAGGCTGACAGTGTGGGCCCTGATGTATGATGATGATGATGATGATGATGATGATG 1399

Qy 1201 ATCGTTAGCTGGGGCTATGCTCGGGGGCCCGAGCACCCAGAGTATACACCAAGGTC 1260
Db 1400 ATCGTTAGCTGGGGCTATGCTCGGGGGCCCGAGCACCCAGAGTATACACCAAGGTC 1459
Qy 1261 TCAGCCTATCTCAACTGGATCTCAATGCTGSAAGGCTGAGCTG 1305
Db 1460 TCAGCCTATCTCAACTGGATCTCAATGCTGSAAGGCTGAGCTG 1504

RESULT 3

BD137129 2038 bp DNA linear PAT 18-SEP-2002
LOCUS Human protease molecule.
DEFINITION
BD137129
ACCESSION
BD137129.1 GI:23232074
VERSION
JP 2002508970-A/6.
KEYWORDS
Homo sapiens (human)
SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 2038)
AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
Tang,T.Y. and Shah,P.
TITLE Human protease molecule
JOURNAL INCYTE PHARMACEUTICALS INC
COMMENT Patent: JP 2002508970-A 6 26-MAR-2002;
OS Homo sapiens (human)
PN JP 2002508970-A/6
PD 26-MAR-2002
PF 12-JAN-1999 JP 2000540252
PR 16-JAN-1998 US 09/008271
PI OLGA BANDMAN,JENNIFER L HILLMAN,HENRY YUE,KARL J GUEGLER,NEIL
PI C CORLEY,
PI TOM Y TANG,PURVI SHAH
PC C12N15/09,A61K38/46,C07K16/40,C12N1/19,C12N1/21,C12N5/10 PC
C12N9/48,C12N9/64,
PC C12Q1/68,C12N15/00,A61K37/54,C12N5/00
CC 1337018
FH Key Location/Qualifiers
FT source 1..2038 /organism='Homo sapiens (human)'.
FT Location/Qualifiers

FEATURES

source 1..2038
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN

Query Match 99.9%; Score 1303.4; DB 6; Length 2038;
Best Local Similarity 99.9%; Pred. No. 1.8e-301;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGGCAACCC 60
Db 200 ATGGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGGCAACCC 259
Qy 61 CGTATCCCCATGAGACCTTCAGAAAGTGGGATCCCCATCATATAGCACTACTGAGC 120
Db 260 CGTATCCCCATGAGACCTTCAGAAAGTGGGATCCCCATCATATAGCACTACTGAGC 319
Qy 121 CTGGCAGATCATCATATTGTGTTGCTTCATCAAGGTGATTCGTGATAAATACTACTTC 180
Db 320 CTGGCAGATCATCATATTGTGTTGCTTCATCAAGGTGATTCGTGATAAATACTACTTC 379
Qy 181 CTCTGGGGCAGCTCTCCACTTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC 240
Db 380 CTCTGGGGCAGCTCTCCACTTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC 439
Qy 241 TGTCCCTTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCACTG 300
Db 440 TGTCCCTTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCACTG 499

QY	301	GCAGTCCGCTCTCCAGGACCGATCCACTGACGAGTGTCTGGACTCGGCCACAGGGAAC	360
DB	500	GCAGTCCGCTCTCCAGGACCGATCCACTGACGAGTGTCTGGACTCGGCCACAGGGAAC	559
QY	361	TGGTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACGCTGTAGGACG	420
DB	560	TGGTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACGCTGTAGGACG	619
QY	421	ATGGCTACAGCAGCAAAACCTTCAGAGCTGTGGAGATGGCCAGACGAGATCTG	480
DB	620	ATGGCTACAGCAGCAAAACCTTCAGAGCTGTGGAGATGGCCAGACGAGATCTG	679
QY	481	GATGTTGTGAATCACAGAAAACGACGAGCTTCGATGCGGAATCAAGTGGGCC	540
DB	680	GATGTTGTGAATCACAGAAAACGACGAGCTTCGATGCGGAATCAAGTGGGCC	739
QY	541	TGCTCTCAGGCTCCCTGCTCTCCCTGACCTGCTGTGCTGTGGAGAGCTTCAAGACC	600
DB	740	TGCTCTCAGGCTCCCTGCTCTCCCTGACCTGCTGTGCTGTGGAGAGCTTCAAGACC	799
QY	601	CCCCGTGTGGTGGGAGAGGCTCTCTGGAATCTTGGCTTGGCAGGTCAAGTCA	660
DB	800	CCCCGTGTGGTGGGAGAGGCTCTCTGGAATCTTGGCTTGGCAGGTCAAGTCA	859
QY	661	CAGTACGACAAACACAGCTCTGTGGAGGAGCATCTGACCCCTGAGTGGTCTCAGG	720
DB	860	CAGTACGACAAACACAGCTCTGTGGAGGAGCATCTGACCCCTGAGTGGTCTCAGG	919
QY	721	GCAGGCCACTGCTTCAGAAAACATACCGATGTTTCACTGGAAGGTGGGCGAGGTCA	780
DB	920	GCAGGCCACTGCTTCAGAAAACATACCGATGTTTCACTGGAAGGTGGGCGAGGTCA	979
QY	781	GACAACTGGGAGCTTCCATCTCCCTGCTGTGGCAGAGTCAATCATTTGAATCAAC	840
DB	980	GACAACTGGGAGCTTCCATCTCCCTGCTGTGGCAGAGTCAATCATTTGAATCAAC	1039
QY	841	CCCATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAATGCCACTCACTTTC	900
DB	1040	CCCATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAATGCCACTCACTTTC	1059
QY	901	TCAGGCACAGTCAGGCCCATCTGCTGCTCCCTTTTGTATGAGGAGCTCACTCCAGCCACC	960
DB	1100	TCAGGCACAGTCAGGCCCATCTGCTGCTCCCTTTTGTATGAGGAGCTCACTCCAGCCACC	1159
QY	961	CCACTCTGATCATTTGGATGGGCTTTACGAGCAGATGAGGAGAGATGCTTGACATA	1020
DB	1160	CCACTCTGATCATTTGGATGGGCTTTACGAGCAGATGAGGAGAGATGCTTGACATA	1219
QY	1021	CTGCTGCAGGCGTCACTCCAGTCAATTCACAGCACACGCTCAATTCAGACGATCCGTAC	1080
DB	1220	CTGCTGCAGGCGTCACTCCAGTCAATTCACAGCACACGCTCAATTCAGACGATCCGTAC	1279
QY	1081	CAGGGGAGTCAACGAGAGATGATGTGAGCAGCATCCGAGAGGGGTGTGGACACC	1140
DB	1280	CAGGGGAGTCAACGAGAGATGATGTGAGCAGCATCCGAGAGGGGTGTGGACACC	1339
QY	1141	TCCACAGGCTGACAGTGGTGGGCTTGTGATGACCAATCTGACCAATGTCATGCTGGTGGG	1200
DB	1340	TCCACAGGCTGACAGTGGTGGGCTTGTGATGACCAATCTGACCAATGTCATGCTGGTGGG	1399
QY	1201	ATCGTTAGTGGGGCTATGCTGCGGGGCGGACACCCAGGAGTATACCAAGTCT	1260
DB	1400	ATCGTTAGTGGGGCTATGCTGCGGGGCGGACACCCAGGAGTATACCAAGTCT	1459
QY	1261	TCAGCTATCTCACTGATCTACATGCTGGAAGGCTGAGCTG	1305
DB	1460	TCAGCTATCTCACTGATCTACATGCTGGAAGGCTGAGCTG	1504

RESULT 4
BC011703
LOCUS
DEFINITION Homo sapiens transmembrane protease, serine 4, transcript variant

ACCESSION	BC011703
VERSION	BC011703.2
KEYWORDS	GI:33991397
SOURCE	MGC.
ORGANISM	Homo sapiens (human)
REFERENCE	1 (bases 1 to 2079) Strausberg, R.L., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, T.E., Brownstein, M.J., Usdin, T.B., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Casavant, T.L., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Kravinsky, M.I., Skalska, J., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
TITLE	2 (bases 1 to 2079) Strausberg, R. Direct submission Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov On Aug 20, 2003 this sequence version replaced gi:15079794. Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-t., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
REMARK	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 27 Row: f Column: 19 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 15451939.
FEATURES	Location/Qualifiers 1..2079 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:19490 IMAGE:3610695" /tissue_type="Pancreas, adenocarcinoma" /clone_lib="NIH_MGC_39"


```
/lab host="DH10B-R"
/notes="vector: pOTB7"
1. 2079
/genes="TMPSR34"
/notes="synonyms: TMPSR3, MT-SP2"
/db_xref="LocusID:56649"
/db_xref="MIM:606565"
208. 1521
/codon_start=1
/product="transmembrane protease, serine 4, isoform 1"
/protein_id="AAH11703.1"
/db_xref="GI:15079795"
/db_xref="LocusID:56649"
/translacion="MLQDPSDQPLNSLDVVKLRKPRIPMETFRKVGIPILIALSLA
SIIIVLILKVIDLYFLQCPHLFIPRQLDGLDCPLGEDEHCKVSPFPGAV
AVRLKRSLOVLDSATGNWFGACFDNTEALAEACROMYGSXKPTFAVIGPQD
DLVVEITENSLRWNSGSLGSLVSLHCLAGKSLKTRPVVGGEEASVDSWPW
QVSTQYKHVCGSILDPHWLTAHCRKHTDVENWKRASDKLGFPSLAVAKI
ILIEFPMYPRNDIALMKLQFLPTFSGRVPLCFDFBELPAPLPMILIGWFTKQ
NGKVSILLQASVQVLDSTFCNADDAVQGEVTERMKAGIPGGVDVTCQDSSGFLM
YQSDQHVGVGIVSNGYCGGPFSTPGVYTRKVSAYLNIWYVWKAEL"
379. 483
/notes="LDLa; Region: Low Density Lipoprotein Receptor
Class A domain, a cysteine-rich repeat that plays a
central role in mammalian cholesterol metabolism"
/db_xref="CDD:cd00112"
517. 789
/notes="SR; Region: Scavenger receptor Cys-rich"
/db_xref="CDD:smart00202"
817. 1494
/notes="Tryp SPc; Region: Trypsin-like serine protease"
/db_xref="CDD:smart00020"

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 2.2e-301; Length 2079;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCACACCTCTGGCAACCCCG 62
Db 216 GGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCACACCTCTGGCAACCCCG 275
Qy 63 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATGACACTACTGAGCCT 122
Db 276 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATGACACTACTGAGCCT 335
Qy 123 GCGAGTATCATATGTTGTTGCTTCATCAAGGTGATTCGATATAAATCTTCTCT 182
Db 336 GCGAGTATCATATGTTGTTGCTTCATCAAGGTGATTCGATATAAATCTTCTCT 395
Qy 183 CTGGGCGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGACGAGCTGGAGCTG 242
Db 396 CTGGGCGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGACGAGCTGGAGCTG 455
Qy 243 TCCCTTGGGGAGGACGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCTTCAGTGGC 302
Db 456 TCCCTTGGGGAGGACGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCTTCAGTGGC 515
Qy 303 AGTCGCTCTCCAGGACCGATCCACTGACGAGTGTGACTCGGCGACAGGAGCTG 362
Db 516 AGTCGCTCTCCAGGACCGATCCACTGACGAGTGTGACTCGGCGACAGGAGCTG 575
Qy 363 GTTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGAGACGCTGTAGGCAGAT 422
Db 576 GTTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGAGACGCTGTAGGCAGAT 635
Qy 423 GGGCTACAGCAGCAAAACCACTTTCAGAGCTGTGAGATTTGGCCACAGCAGATCTGGA 482
Db 636 GGGCTACAGCAGCAAAACCACTTTCAGAGCTGTGAGATTTGGCCACAGCAGATCTGGA 695
Qy 483 TGTGTTGAATACAGAAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCTG 542
Db 696 TGTGTTGAATACAGAAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCTG 755
```

```
Qy 543 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTCAAGACCCC 602
Db 756 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTCAAGACCCC 815
Qy 603 CCGTGTGTGGTGGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGAGATCCA 662
Db 816 CCGTGTGTGGTGGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGAGATCCA 875
Qy 663 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACACCCCTCTGGTCTCTCACGGC 722
Db 876 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACACCCCTCTGGTCTCTCACGGC 935
Qy 723 AGCCCACTCTCTTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGTCTAGA 782
Db 936 AGCCCACTCTCTTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGTCTAGA 995
Qy 783 CAAACTGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTAATTCACACC 842
Db 996 CAAACTGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTAATTCACACC 1055
Qy 843 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAAGTTCGCCACTCACTTCTC 902
Db 1056 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAAGTTCGCCACTCACTTCTC 1115
Qy 903 AGCAGAGTCAGCCCATCTCTGTCTGCTTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 962
Db 1116 AGCAGAGTCAGCCCATCTCTGTCTGCTTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 1175
Qy 963 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAGTGAAGGAGGATGCTGTGATATCT 1022
Db 1176 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAGTGAAGGAGGATGCTGTGATATCT 1235
Qy 1023 GTTCAGGCGTCAGTCCAGGTCAATGACAGACACAGGTGCAATGACAGATGCGTATCCA 1082
Db 1236 GTTCAGGCGTCAGTCCAGGTCAATGACAGACACAGGTGCAATGACAGATGCGTATCCA 1295
Qy 1083 GGGGGAAGTCACGAGAAATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACCTG 1142
Db 1296 GGGGGAAGTCACGAGAAATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACCTG 1355
Qy 1143 CCAGGCTGACAGTGTGGGCTTGTGATGTACCAATGTGACAGTGGCATGTGGTGGCAT 1202
Db 1356 CCAGGCTGACAGTGTGGGCTTGTGATGTACCAATGTGACAGTGGCATGTGGTGGCAT 1415
Qy 1203 CGTTAGCTGGGCTATGGCTGCGGGGCGGAGCAGCAGGATATACCAAGGTCTC 1262
Db 1416 CGTTAGCTGGGCTATGGCTGCGGGGCGGAGCAGCAGGATATACCAAGGTCTC 1475
Qy 1263 AGCTATCTCAACTGATCTACAATGTCTGGAAGGTGAGCTG 1305
Db 1476 AGCTATCTCAACTGATCTACAATGTCTGGAAGGTGAGCTG 1518

RESULT 5
AF179224 2081 bp mRNA linear PRI 08-JUN-2000
LOCUS Homo sapiens transmembrane serine protease 3 (TMPSR3) mRNA,
DEFINITION complete cds.
ACCESSION AF179224
VERSION AF179224.1 GI:8347148
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 2081)
JOURNAL Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
MEDLINE Iwamura,T., Rutenburger,M., Lerch,M.N., Adler,G. and Gress,T.M.
A novel transmembrane serine protease (TMPSR3) overexpressed in
pancreatic cancer
Cancer Res. 60 (10), 2602-2606 (2000)
20283276
```

10825129	PUBMED	2 (bases 1 to 2081)	QY	63	TATCCCCATGGAGACCTTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGCCT	122
REFERENCE	Wallrapp, C. and Gress, T.M.	Direct Submission	DB	283	TATCCCCATGGAGACCTTTCAGAAAGGTGGGATCCCCATCATCATAGCACTACTGAGCCT	342
AUTHORS			QY	123	GGCGAGTATCATCAITGTGTTGTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT	182
JOURNAL	Submitted (20-AUG-1999) Internal Medicine I, University of Ulm, Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany	Location/Qualifiers	DB	343	GGCGAGTATCATCAITGTGTTGTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT	402
FEATURES			QY	183	CTGGGGAGCCTTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG	242
source	1..2081	/organism="Homo sapiens"	DB	403	CTGGGGAGCCTTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG	462
	/mol_type="mRNA"	/db_xref="taxon:9606"	QY	243	TCCCTTGGGGAGGACGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC	302
	/db_xref="taxon:9606"	/chromosome="11"	DB	463	TCCCTTGGGGAGGACGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC	522
	/map="11q23.3"	/tissue_type="pancreatic carcinoma"	QY	303	AGTCCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGACTGGCCACAGGAACTG	362
gene	1..2081	/product="transmembrane serine protease 3"	DB	523	AGTCCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGACTGGCCACAGGAACTG	582
CDS	215..1528	/protein_id="AAF74526.1"	QY	363	GTTCCTGCTCTTCGCAAACTTTCAGAGAGCTCTCGCTGAGAGAGCTGTAGGACAGT	422
	/db_xref="GI:8347149"	/translation="MLQPDSDQPLNSLDVPLKPRIPMETPRKVGIPILIALLSIA SIIIVLILKILDKYFLQCPQLHIFPRKQLCDGELDCPLGDEEHCHVKSFPFEGAV AVRLSKSTLQVLDSATGNWFSACFDNFTEALAEACRQMGYSKFTPEAIGPDQ DLIVVEITENSQELMRNSGFLSGSLVSLHCLACGKSLKTRPVGGEASVDSWPM QVSYQDKOHPYGGGSLDPLHVLTAACHFRKHTDFVFNKVRAGSKLGSFPLAVAKI IIEFNPMYKNDNIALMKLQPLTFSGTVPICLPFDEBELPATPLIWMGFTKQ NGKXNSDILLOASVQVDSIRCNADDAQSEVTEKMCAGIPGSGVDTQGDSDGGLM YQDQMHVGVISVGYCGGSPTEGVITKVSAYLNWIYNWKAEL"	DB	583	GTTCCTGCTCTTCGCAAACTTTCAGAGAGCTCTCGCTGAGAGAGCTGTAGGACAGT	642
misc_feature	308..373	/gene="TMPRSS3"	QY	423	GGGCTACAGCAAAACCCACTTTTTCAGAGCTGTGAGATTTGGCCACAGGATCTTGG	482
	/note="Region: cytoplasmic domain"		DB	643	GGGCTACAGCAAAACCCACTTTTTCAGAGCTGTGAGATTTGGCCACAGGATCTTGG	702
misc_feature	374..1525	/gene="TMPRSS3"	QY	483	TGTTGTGAAATTCAGAAAACAGCAGGAGCTTCCGATCGGAGCTCAAGTGGSCCTG	542
	/note="Region: transmembrane domain"		DB	703	TGTTGTGAAATTCAGAAAACAGCAGGAGCTTCCGATCGGAGCTCAAGTGGSCCTG	762
misc_feature	602..604	/gene="TMPRSS3"	QY	543	TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGAGAGCCTGAAGACCC	602
	/note="Region: extracellular domain"		DB	763	TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGAGAGCCTGAAGACCC	822
misc_feature	746..748	/gene="TMPRSS3"	QY	603	CCGTGTGGTGGTGGGAGGAGGCTCTGTGAGATTTTGGCTTGGCAGGTGAGGATCCA	662
	/note="glycosylation site"		DB	823	CCGTGTGGTGGTGGGAGGAGGCTCTGTGAGATTTTGGCTTGGCAGGTGAGGATCCA	882
misc_feature	947..949	/gene="TMPRSS3"	QY	663	GTACGACAAAACAGCAGCTCTGTGGAGGAGGAGCTCTGTGAGATTTTGGCTTGGC	722
	/note="disulfide bond"		DB	883	GTACGACAAAACAGCAGCTCTGTGGAGGAGGAGCTCTGTGAGATTTTGGCTTGGC	942
misc_feature	1082..1084	/gene="TMPRSS3"	QY	723	AGCCCACTCTTCAGAAAACATACCAGTGTGTTAACTGGAAGGTGCGGGCAGGCTCAG	782
	/note="disulfide bond"		DB	943	AGCCCACTCTTCAGAAAACATACCAGTGTGTTAACTGGAAGGTGCGGGCAGGCTCAG	1002
misc_feature	1373..1375	/gene="TMPRSS3"	QY	783	CAAACTGGGACGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACCC	842
	/note="active site"		DB	1003	CAAACTGGGACGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACCC	1062
misc_feature	1528..1528	/gene="TMPRSS3"	QY	843	CATGTACCCCAAGAACATATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTCTC	902
	/note="active site"		DB	1063	CATGTACCCCAAGAACATATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTCTC	1122
misc_feature	1528..1528	/gene="TMPRSS3"	QY	903	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTCATGAGGAGCTCACTTCAGGCCACCC	962
	/note="disulfide bond"		DB	1123	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTCATGAGGAGCTCACTTCAGGCCACCC	1182
misc_feature	1528..1528	/gene="TMPRSS3"	QY	963	ACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGAAATGGAGGAGATGTCTGACATACT	1022
	/note="active site"		DB	1183	ACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGAAATGGAGGAGATGTCTGACATACT	1242
ORIGIN			QY	1023	GTGTGAGGCTCAGTCCAGGTCATTTGACAGACAGGTCGATGCGAGAGATGGGTACCA	1082
			DB	1243	GTGTGAGGCTCAGTCCAGGTCATTTGACAGACAGGTCGATGCGAGAGATGGGTACCA	1302
			QY	1083	GGGGGAAGTCAACCGAAGATGATGTGTGAGGATTCGCCGGAAGGGGTGTGGACACCTG	1142
			DB	1303	GGGGGAAGTCAACCGAAGATGATGTGTGAGGATTCGCCGGAAGGGGTGTGGACACCTG	1362
			QY	1143	CCAGGGTACAGTGTGGGCCCCCTGATGTACCACTCTGACACGAGTGTGGTGGCAT	1202

Query Match 99.8%; Score 1303; DB 9; Length 2081;

Best Local Similarity 100.0%; Pred. No. 2.2e-301;

Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTTGACAGTATCAACCTCTGACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 62

DB 223 GGAATCTTGACAGTATCAACCTCTGACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 282

```

Db      1363  CCAGGGTGCACAGTGGTGGGCCCTCGATGTACCAATCTGACAGTGGCATGTGTGGGCAT 1422
Qy      1203  CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCACAGGAGTATACACCAAGGTCTC 1262
Db      1423  CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCACAGGAGTATACACCAAGGTCTC 1482
Qy      1263  AGCCTATCTCAACTGGATCTACAACTGTCTGGAAGGCTGAGCTG 1305
Db      1483  AGCCTATCTCAACTGGATCTACAACTGTCTGGAAGGCTGAGCTG 1525

RESULT 6
AX207899 LOCUS 2137 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0157194.
ACCESSION AX207899
VERSION AX207899.1 GI:15422497
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Madison,E.L., Ong,E.O. and Yeh,J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
JOURNAL encoded proteins and methods based thereon
PARENT: WO 0157194-A 3 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source Location/Qualifiers
1. .2137
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
261. .1574
/notes="unnamed protein product; DNA sequence encoding a
transmembrane serine protease (MtsP3) protein"
/codon_start=1
/protein_id="CAC60378.1"
/db_xref="GI:15422498"
/db_xref="REMBL:CA60378"
/translation="MLQPPSDQVPLNSLDVPLKPKRIPMETRKKVGIPIIIALLSLA
SIIITVVLKIDYFLGQQLHFIIPKQCDGELDCPLGDEEHCVKSPFEGFAV
AVRLSKRSTLQVLDSTGNMFSACFDNTEALAEACRQMYSSKFTFPAVIGPDQ
LDVVEITENSOELRMENSGPCLSLGSLHCLACGSLKTRPVVGGEEASVDSPW
QVS:QYDIHQVGGGSLDHPWLVTAACHFRKHTDVFNKVRAGSKLGSFPFLAVAKI
IILFEPMYKQNDIALMKLOPLTFSTVRLICLPFDDELTPATPLWLIIGWFTQ
NGKMSDILLOASVQVVDSTRCNADDAVQSEVTEKMKCAGIPGGVDTCCGDSGGPLM
YQSDQWHVVGIVSGVGGGSPTEPVITIKVSAYLNWYNVKAEL"

ORIGIN
Query Match 99.5%; Score 1298.2; DB 6; Length 2137;
Best Local Similarity 99.8%; Pred. No. 3.1e-300;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 GGATCCTGACAGTATCAACTCTGAAACAGCCTGATGTCAAACCCCTGCGCAACCCCG 62
Db 269 GGATCCTGACAGTATCAACTCTGAAACAGCCTGATGTCAAACCCCTGCGCAACCCCG 328
Qy 63 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 329 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 388
Qy 123 GCGAGTATCATATTGGTGTCTCATCAAGGTGATCTGGAATAACTACTTCTCT 182
Db 389 GCGAGTATCATATTGGTGTCTCATCAAGGTGATCTGGAATAACTACTTCTCT 448
Qy 183 CTGCGGGAGCCTCTCCACTTCATCCGAGAGCAGCTGTGACCGAGAGCTGAGCTG 242
Db 449 CTGCGGGAGCCTCTCCACTTCATCCGAGAGCAGCTGTGACCGAGAGCTGAGCTG 508
Qy 243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCTGCACTGGC 302
```

RESULT 7
AX076192

```

Db      509  TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGC 568
Qy      303  AGTCGCGCTTCTCAAGGACCGATCCACTGTGAGGTGTGTGACTCGGCCACAGGGAACCTG 362
Db      569  AGTCGCGCTTCTCAAGGACCGATCCACTGTGAGGTGTGTGACTCGGCCACAGGGAACCTG 628
Qy      363  GTTCTCTGCTGCTTTTCGACAACTTCCACAGAAGCTCTCGCTGAGACAGCCTGTAGCAGAT 422
Db      629  GTTCTCTGCTGCTTTTCGACAACTTCCACAGAAGCTCTCGCTGAGACAGCCTGTAGCAGAT 688
Qy      423  GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGAGATTTGGCCCGACAGCAGATCTGGA 482
Db      689  GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGAGATTTGGCCCGACAGCAGATCTGGA 748
Qy      483  TGTGTGTGAATTCACAGAAACAGCCAGAGCTTCGCAATGCGGAATCAAGTGGGCCCTG 542
Db      749  TGTGTGTGAATTCACAGAAACAGCCAGAGCTTCGCAATGCGGAATCAAGTGGGCCCTG 808
Qy      543  TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGGAAAGAGCTGGAAGACCCC 602
Db      809  TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGGAAAGAGCTGGAAGACCCC 868
Qy      603  CCGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCCCTTGGCAGGTGAGCATCCA 662
Db      869  CCGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCCCTTGGCAGGTGAGCATCCA 928
Qy      663  GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTGTGACCCCACTGGGTCTCTCACGGC 722
Db      929  GTACGACATACAGCAGCTGTGTGGAGGAGCATCTGTGACCCCACTGGGTCTCTCACGGC 988
Qy      723  AGCCCACTGCTTCAGGAAACATACCGATGTGTTCATCTGGAAGGTGGGGAGGCTCAGA 782
Db      989  AGCCCACTGCTTCAGGAAACATACCGATGTGTTCATCTGGAAGGTGGGGAGGCTCAGA 1048
Qy      783  CAACTGGGAGCTTCCCATCTGCTGTGCGCAAGATCATCATCATCATCATCATCATCATCAT 842
Db      1049  CAACTGGGAGCTTCCCATCTGCTGTGCGCAAGATCATCATCATCATCATCATCATCATCAT 1108
Qy      843  CATGTACCCCAAGCAATGACATCGCCCTCATAGAGCTGCAAGTTCCTCACTCACTTTCTC 902
Db      1109  CATGTACCCCAAGCAATGACATCGCCCTCATAGAGCTGCAAGTTCCTCACTCACTTTCTC 1168
Qy      903  AGGCACAGTCAGCCCATCTGTCTGCGCTTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 962
Db      1169  AGGCACAGTCAGGCTCATCTGTCTGCGCTTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 1228
Qy      963  ACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGAGGAGATGTCTGACATACT 1022
Db      1229  ACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGAGGAGATGTCTGACATACT 1288
Qy      1023  GCTGCGAGGCGTCAAGTCCAGTCATTTGACAGCACCGGTGCAATGACAGATGCGGTACCA 1082
Db      1289  GCTGCGAGGCGTCAAGTCCAGTTCATTTGACAGCACCGGTGCAATGACAGATGCGGTACCA 1348
Qy      1083  GGGGGAAGTCACCGAAGATGATGTGTCAGGATCCCGGAAGGGGGTGTGGACACCTG 1142
Db      1349  GGGGGAAGTCACCGAAGATGATGTGTCAGGATCCCGGAAGGGGGTGTGGACACCTG 1408
Qy      1143  CCAGGTGACAGTGGTGGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1202
Db      1409  CCAGGTGACAGTGGTGGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1468
Qy      1203  CGTGTAGTGGGCTATGGCTGGGGGGCCCGAGCAGCCCGAGGAGTATACACCAAGGTCTC 1262
Db      1469  CGTGTAGTGGGCTATGGCTGGGGGGCCCGAGCAGCCCGAGGAGTATACACCAAGGTCTC 1528
Qy      1263  AGCCTATCTCAACTGGATCTTACAACTGTCTGGAAGGCTGAGCTG 1305
Db      1529  AGCCTATCTCAACTGGATCTTACAACTGTCTGGAAGGCTGAGCTG 1571
```


DIADIXUS INC
OS Homo sapiens (human)
FN JP 2002523760-A/2
PD 30-JUL-2002
PF 01-SEP-1999 JP 2000567741
PR 02-SEP-1998 US 60/098880
PI SUSANA SALCEDA, YONGXING SUN, HERVE RECIPON, ROBERT CAFFERKEY PC
GOIN33/574, A61K39/395, A61K39/395, A61K39/395, A61K39/395 PC
, A61K49/00, A61K49/00,
PC A61K51/00, C07K16/32, C12N15/09, C12Q1/68, C12N15/00, A61K49/02 CC
A novel method of diagnosing, monitoring,
staging, imaging and
treating
CC various cancers
FH Key Location/Qualifiers
FT source 1. .2070
FT source /organism="Homo sapiens (human)"
FT source Location/Qualifiers
1. .2070
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source

ORIGIN

Query Match 99.0%; Score 1292; DB 6; Length 2070;
Best Local Similarity 99.9%; Pred. No. 9.5e-299;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCCCG 62
DB 222 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCCCG 281
63 TATCCCTGGAGAGCCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
DB 282 TATCCCTGGAGAGCCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 341
123 GCGAGTATCATATGTTGGTTCCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 182
DB 342 GCGAGTATCATATGTTGGTTCCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 401
183 CTGGGGGAGCCTCCACTTCATCCGAGGAGCAGCTGTGTACGAGAGCTGGACTG 242
DB 402 CTGGGGGAGCCTCCACTTCATCCGAGGAGCAGCTGTGTACGAGAGCTGGACTG 461
243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGC 302
DB 462 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGC 521
303 AGTCGGCTCTCCAGGACCGATCCACACTGAGTGTGACTCGGCTCGGCCACAGGGAATG 362
DB 522 AGTCGGCTCTCCAGGACCGATCCACACTGAGTGTGACTCGGCTCGGCCACAGGGAATG 581
363 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTGCTGAGACAGCTCTAGGCAGAT 422
DB 582 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTGCTGAGACAGCTCTAGGCAGAT 641
423 GGGCTACAGAGCAAAACCACTTTTCAGAGCTGTGAGATTGGCCAGACAGGATCTGGA 482
DB 642 GGGCTACAGAGCAAAACCACTTTTCAGAGCTGTGAGATTGGCCAGACAGGATCTGGA 701
483 TGTGTTCAATTCACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCCCTG 542
DB 702 TGTGTTCAATTCACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCCCTG 761
543 TCTCTCAGGCTCCCTGGTCTCCCTGCATGTTCTTGCTGTGGGAGAGCCTGAAGACCCC 602
DB 762 TCTCTCAGGCTCCCTGGTCTCCCTGCATGTTCTTGCTGTGGGAGAGCCTGAAGACCCC 821
603 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGTGCAGATCCA 662
DB 822 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGTGCAGATCCA 881
663 GTACGACAAAAGCAGACGCTCTGTGGAGGAGGATCCTGTGACCCCACTGGGTCTTCAC-GG 721

DB 882 GTACGACAAAAGCAGACGCTCTGTGAGGAGAGATCTCTGGACCCCTCTCTCAGCGG 941
QY 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGCGAGGCTCAG 781
DB 942 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGCGAGGCTCAG 1001
QY 782 ACAAATGGGAGCTTCCCATCCCTGGCTGTGCCCAGAGATCATCATCAATTGAATCAACC 841
DB 1002 ACAAATGGGAGCTTCCCATCCCTGGCTGTGCCCAGAGATCATCATCAATTGAATCAACC 1061
QY 842 CCATGTACCCCAAGACAAATGACATGCCCTCATGAGCTGCAGTTCGCCACTCATTCT 901
DB 1062 CCATGTACCCCAAGACAAATGACATGCCCTCATGAGCTGCAGTTCGCCACTCATTCT 1121
QY 902 CAGGACAGTCCAGCCCATCTGCTGCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 961
DB 1122 CAGGACAGTCCAGCCCATCTGCTGCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1181
QY 962 CACTCTGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAGAGATGCTTGACATAC 1021
DB 1182 CACTCTGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAGAGATGCTTGACATAC 1241
QY 1022 TGCTGAGGCTCAGTCCAGTCAATTCAGACACACGCTCAATGCAGAGATGCGTACC 1081
DB 1242 TGCTGAGGCTCAGTCCAGTCAATTCAGACACACGCTCAATGCAGAGATGCGTACC 1301
QY 1082 AGGGGAGTCCAGGAGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACT 1141
DB 1302 AGGGGAGTCCAGGAGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACT 1361
QY 1142 GCCAGGTCAGTGTGGGCTGATGTACCAATCTGACCACTGGCTGGTGGCA 1201
DB 1362 GCCAGGTCAGTGTGGGCTGATGTACCAATCTGACCACTGGCTGGTGGCA 1421
QY 1202 TCCTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCAGAGATATACCAAGTCT 1261
DB 1422 TCCTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCAGAGATATACCAAGTCT 1481
QY 1262 CAGCTATCTCACTGGATCAATGTCTGGAAGCTGAGCTG 1305
DB 1482 CAGCTATCTCACTGGATCAATGTCTGGAAGCTGAGCTG 1525

RESULT 9
AR232520
LOCUS AR232520 2079 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6455668.
ACCESSION AR232520
VERSION AR232520.1 GI:27274657
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2079)
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.
TITLE Methods of diagnosing colorectal cancer, compositions, and methods of screening for colorectal cancer modulators
JOURNAL Patent: US 6455668-A 1 24-SEP-2002;
FEATURES Location/Qualifiers
source 1. .2079
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 99.0%; Score 1292; DB 6; Length 2079;
Best Local Similarity 99.9%; Pred. No. 9.5e-299;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCCCG 62
DB 216 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCGAAACCCCG 275

63 TATCCCATGGAGACCTTCAGAAAG-GTGGGGATCCCATCATCATAGCACTACTGAGCC 121
DB
276 TATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY
122 TGGCAGATATCATCATTTGTTGCTCTCATCAAGGTGATTCGATGAATAACTACTTCC 181
DB
336 TGGCAGATATCATCATTTGTTGCTCTCATCAAGGTGATTCGATGAATAACTACTTCC 395
QY
182 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGCACT 241
DB
396 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGCACT 455
QY
242 GTCCCTTGGGGAGGACGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGATGG 301
DB
456 GTCCCTTGGGGAGGACGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGATGG 515
QY
302 CAGTCCGCTCTCCAGAGCAGTCCACACTGCAAGTGTGAGTCCGCGACAGGGAAT 361
DB
516 CAGTCCGCTCTCCAGAGCAGTCCACACTGCAAGTGTGAGTCCGCGACAGGGAAT 575
QY
362 GGTCTCTGCTCTGTTTGGACAACTTCAGAAAGCTTCCGCTGAGACAGCTGTAGGAGA 421
DB
576 GGTCTCTGCTCTGTTTGGACAACTTCAGAAAGCTTCCGCTGAGACAGCTGTAGGAGA 635
QY
422 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTTGGCCGACAGGATCTGG 481
DB
636 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTTGGCCGACAGGATCTGG 695
QY
482 ATGTGTTTGAATTCAGAAACACCCAGGAGCTTCGATCGGAATTCAGTGGGCGCT 541
DB
696 ATGTGTTTGAATTCAGAAACACCCAGGAGCTTCGATCGGAATTCAGTGGGCGCT 755
QY
542 GTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGAGAGAGCTGTAGACCC 601
DB
756 GTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGAGAGAGCTGTAGACCC 815
QY
602 CCGGTGTGTGTGGTGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCAGGTCAAGCATCC 661
DB
816 CCGGTGTGTGTGGTGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCAGGTCAAGCATCC 875
QY
662 AGTAGCAAAACAGACAGCTCTGTGAGGAGAGCATCTCGAACCCCATCTGGTCTCTCAGG 721
DB
876 AGTAGCAAAACAGACAGCTCTGTGAGGAGAGCATCTCGAACCCCATCTGGTCTCTCAGG 935
QY
722 CAGCCCACTGCTTCAGGAACATACCGATGTTTCAATGGAAGTGGGCGAGCTCAG 781
DB
936 CAGCCCACTGCTTCAGGAACATACCGATGTTTCAATGGAAGTGGGCGAGCTCAG 995
QY
782 ACAAACTGGGCGAGCTTCCCATCCCTGGTGTGGCCAAAGATCATCATTTGAATCAACC 841
DB
996 ACAAACTGGGCGAGCTTCCCATCCCTGGTGTGGCCAAAGATCATCATTTGAATCAACC 1055
QY
842 CATATGACCCCAAGAACATATGATCGCTCATGAAGCTGAGTCCCATCTCACTTCT 901
DB
1056 CATATGACCCCAAGAACATATGATCGCTCATGAAGCTGAGTCCCATCTCACTTCT 1115
QY
902 CAGGCACTGAGGCGCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
DB
1116 CAGGCACTGAGGCGCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
QY
962 CACTGTGATCATTTGATGGGCTTTACGAACAGAAATGGAGGAGAGATGCTGACATAC 1021
DB
1176 CACTGTGATCATTTGATGGGCTTTACGAACAGAAATGGAGGAGAGATGCTGACATAC 1235
QY
1022 TGTGAGGCGCTCAGTTCAGAGTCAATGACAGACACCGTTCGAATGACAGAGATGGGTACC 1081
DB
1236 TGTGAGGCGCTCAGTTCAGAGTCAATGACAGACACCGTTCGAATGACAGAGATGGGTACC 1295
QY
1082 AGGGGAAGTACCCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1141
DB
1296 AGGGGAAGTACCCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1355
QY
1142 GCCAGGCTGACAGTGTGGGCGCTTGTATGATCAATCTGACAGTGGCATGTGGTGGCA 1201

1356 GCAGGGTACAGTGTGGGCCCCCTGATGTACCAATCTGACCAATCTGACCAATCTGGTGGCA 1415
QY
1202 TCGTTAGCTGGGCGCTATGGCTCGGGGGCCCCGAGCACCCAGAGTATACACCAAGTCT 1261
DB
1416 TCGTTAGCTGGGCGCTATGGCTCGGGGGCCCCGAGCACCCAGAGTATACACCAAGTCT 1475
QY
1262 CAGCCTATCTCACTGAGTCAATGCTCAATGCTGGAAGCTGAGCTG 1305
DB
1476 CAGCCTATCTCACTGAGTCAATGCTGGAAGCTGAGCTG 1519
RESULT 10
AX207967 2079 bp DNA linear PAT 31-AUG-2001
LOCUS Sequence 71 from Patent WO0157194.
DEFINITION AX207967
ACCESSION AX207967
VERSION AX207967.1 GI:15422563
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 71 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
Location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
251..1522
/note="unnamed protein product; Nucleotide sequence encoding transmembrane protease, serine 4 (TPRSS4)"
/codon_start=1
/protein_id="C660389.1"
/db_xref="GI:15422564"
/db_xref="EMBL:AC60389"
/translation="MSNPANPVSRRPSESIGPIIIALLSLASIIIVLVILKVID
KYFLCGQLHPIPRKQCDGELDPLGDEHCVKSPSPGPAVAVRLSKORSTLOVL
DSATGNFSAFDNFTALAEACQKQYSSFTFRVAVETGQDLVDVEITENSGEL
RWNSGPGCLSGSLVSLHCLACSKLTPRVYGGEEASVDSPWQVSIQYDKHVCQG
SILDPHWLTAACHFRKHTDVFNVKVRAGSKLGSFPPLAVAKIIIFENPMYKDN
IALMKQPLTFTSGTVPICLPFDEBELTPATLWIGWFTKQNGKMSDILLQASV
QVLDTRCNADDAVOGEVTEKMKCAGIPGGVDTCQDSDGGLPMYQSDQWHVVGVSW
GYGCGSPSTPGVYTKVSALNININYNWKAEL"
ORIGIN
Query Match 99.0%; Score 1292; DB 6; Length 2079;
Best Local Similarity 99.9%; Pred. No. 9.5e-299;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCAAACCCCTCGGAAACCCCG 62
DB 216 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCAAACCCCTCGGAAACCCCG 275
QY 63 TATCCCCCATGGAGACCTTCAGAAAG-GTGGGGATCCCATCATCATAGCACTACTGAGCC 121
DB 276 TATCCCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY 122 TGGCAGATATCATCATTTGTTGCTCTCATCAAGGTGATTCGATGAATAACTACTTCC 181
DB 336 TGGCAGATATCATCATTTGTTGCTCTCATCAAGGTGATTCGATGAATAACTACTTCC 395
QY 182 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGCACT 241
DB 396 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGCACT 455
QY 242 GTCCCTTGGGGAGGACGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGATGG 301

QY 422 TGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG 481
DB 536 TGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG 695
QY 482 ATGTGTTGAAATCACAGAAACAGCAGAGCTTCGCATGCGGAATCAAGTGGGCGCT 541
DB 696 ATGTGTTGAAATCACAGAAACAGCAGAGCTTCGCATGCGGAATCAAGTGGGCGCT 755
QY 542 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTCTGTGGAAAGAGCTCAAGACCC 601
DB 756 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTCTGTGGAAAGAGCTCAAGACCC 815
QY 602 CCCTGTGTGGGTGGGAGAGAGCCCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCC 661
DB 816 CCCTGTGTGGGTGGGAGAGAGCCCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCC 875
QY 662 AGTAGCAAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCAAGG 721
DB 876 AGTAGCAAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCAAGG 935
QY 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAG 781
DB 936 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAG 995
QY 782 ACAAACTGGGAGCTTCCCATCTCCCTGCTGTGGCAAGATCATCATATTGAATTCGAAC 841
DB 996 ACAAACTGGGAGCTTCCCATCTCCCTGCTGTGGCAAGATCATCATATTGAATTCGAAC 1055
QY 842 CCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCGCCACTCACTTTCT 901
DB 1056 CCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCGCCACTCACTTTCT 1115
QY 902 CAGGACAGTCAAGCCCACTGTGTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
DB 1116 CAGGACAGTCAAGCCCACTGTGTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
QY 962 CACTCTGATCATTTGATGGGCTTTACGAAGCAGATGAGGGAAGATGCTGACATAC 1021
DB 1176 CACTCTGATCATTTGATGGGCTTTACGAAGCAGATGAGGGAAGATGCTGACATAC 1235
QY 1022 TGCTGCAGGCTCAGTCCAGTCAATTGACACCAACCGGTGCAATGACAGATGCGTACC 1081
DB 1236 TGCTGCAGGCTCAGTCCAGTCAATTGACACCAACCGGTGCAATGACAGATGCGTACC 1295
QY 1082 AGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGGTGTGACACCT 1141
DB 1296 AGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGGTGTGACACCT 1355
QY 1142 GCCAGGCTGACAGTGTGGGCGCCCTGTATGATACCAATCTGACAGTGCATGTGGGCA 1201
DB 1356 GCCAGGCTGACAGTGTGGGCGCCCTGTATGATACCAATCTGACAGTGCATGTGGGCA 1415
QY 1202 TCGTTAGTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1261
DB 1416 TCGTTAGTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1475
QY 1262 CAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1305
DB 1476 CAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1519

RESULT 12
AX092380
LOCUS AX092380 2063 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 111 from Patent WO0116318.
ACCESSION AX092380
VERSION AX092380.1 GI:13444504
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.B., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 111 08-MAR-2001;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
source 1..2063
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 96.8%; Score 1263; DB 6; Length 2063;
Best Local Similarity 98.8%; Pred. NO. 8.4e-292;
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
DB 218 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 277
QY 63 TATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 122
DB 278 TATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 337
QY 123 GCGAGTATCATATGTGGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 182
DB 338 GCGAGTATCATATGTGGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 397
QY 183 CTGCGGGGAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGTGGATG 242
DB 398 CTGCGGGGAGCCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGTGGATG 457
QY 243 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGCCTGAGTGGC 302
DB 458 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGCCTGAGTGGC 517
QY 303 AGTCCGCTCTCCAAAGGACCGATCCACACTGACAGCTGTGGACTCGGCCACAGGGAATG 362
DB 518 AGTCCGCTCTCCAAAGGACCGATCCACACTGACAGCTGTGGACTCGGCCACAGGGAATG 577
QY 363 GTCTCTGCTCTTTTCGACAACTTCACAGAGCTCTCTCTGAGACAGCTGTAGGACAT 422
DB 578 GTCTCTGCTCTTTTCGACAACTTCACAGAGCTCTCTCTGAGACAGCTGTAGGACAT 637
QY 423 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTTGGCCACAGACCAAGATCTGGA 482
DB 638 GGGCTACAG-----CAGAGCTGTGAGATTTGGCCACAGACCAAGATCTGGA 682
QY 483 TGTGTGTAATCACAGAAACAGCCAGAGCTTCGATTCGGAATCTCAAGTGGGCGCTG 542
DB 683 TGTGTGTAATCACAGAAACAGCCAGAGCTTCGATTCGGAATCTCAAGTGGGCGCTG 742
QY 543 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGGAAAGAGCTTGAAGACCCC 602
DB 743 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGGAAAGAGCTTGAAGACCCC 802
QY 603 CCGTGTGTGGGTGGGAGGAGGCGCTCTGTGATTTTGGCTTGGCAGGTGAGATCA 662
DB 803 CCGTGTGTGGGTGGGAGGAGGCGCTCTGTGATTTTGGCTTGGCAGGTGAGATCA 862
QY 663 GTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCAGCGC 722
DB 863 GTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCAGCGC 922
QY 723 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGA 782
DB 923 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGA 982
QY 783 CAAACTGGGAGCTTCCCATCTCCCTGCTGTGGCAGATCATCATGATTAATCAACCC 842
DB 983 CAAACTGGGAGCTTCCCATCTCCCTGCTGTGGCAGATCATCATGATTAATCAACCC 1042

QY 843 CATGTACCCAAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 902
Db 1043 CATGTACCCAAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 1102
QY 903 AGGCACAGTCAGGCCCACTCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCC 962
Db 1103 AGGCACAGTCAGGCCCACTCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1162
QY 963 ACTCTGATCATTTGATGGGCTTTTGAAGAGATGAGGAGGAGTGTCTGACATACT 1022
Db 1163 ACTCTGATCATTTGATGGGCTTTTGAAGAGATGAGGAGGAGTGTCTGACATACT 1222
QY 1023 GCTCAGGCGTCAGTCCAGGTCATTGACAGCACACAGCGTGAATGACAGCATGGTACCA 1082
Db 1223 GCTCAGGCGTCAGTCCAGGTCATTGACAGCACACAGCGTGAATGACAGCATGGTACCA 1282
QY 1083 GGGGAAAGTCACCGAAGAGATGATGTGTGAGGAGTCCCGAAGGGGGTGTGACACCTG 1142
Db 1283 GGGGAAAGTCACCGAAGAGATGATGTGTGAGGAGTCCCGAAGGGGGTGTGACACCTG 1342
QY 1143 CCAGGTCAGAGTGGTGGGCGCTGATGATGACAGTGTACCAATCTGACAGTGGGTCAT 1202
Db 1343 CCAGGTCAGAGTGGTGGGCGCTGATGATGACAGTGTGATGAGTGGGTCAT 1402
QY 1203 CGTTAGCTGGGCTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTC 1262
Db 1403 CGTTAGCTGGGCTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTC 1462
QY 1263 AGCCTATCTCAACTGGATCTACATGTCGAGGCTGAGCTG 1305
Db 1463 AGCCTATCTCAACTGGATCTACATGTCGAGGCTGAGCTG 1505

RESULT 13
AX376262 2063 bp DNA linear PAT 01-MAR-2002
LOCUS
DEFINITION Sequence 329 from Patent WO0168848.
ACCESSION AX376262
VERSION AX376262.1 GI:19170522
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Baker, K.P., Chen, J., Desnoyers, L., Goddard, A., Godowski, P.J., Gurney, A.L., Pan, J., Smith, V., Watanabe, C.K., Wood, W.I. and Zhang, Z.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding the same
JOURNAL Patent: WO 0168848-A 329 20-SEP-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..2063
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 96.8%; Score 1263; DB 6; Length 2063;
Best Local Similarity 98.8%; Pred. No. 8,4e-252;
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 3 GGATCCTGACAGTGATCAACTCTGAAACAGCTCGATGTCAAAACCCCTGGCGAAACCCCG 62
Db 218 GGATCCTGACAGTGATCAACTCTGAAACAGCTCGATGTCAAAACCCCTGGCGAAACCCCG 277
QY 63 TATCCCATGGACCTTCAGAAAGTGGGATCCCATCATCATGACATCTAGCCT 122
Db 278 TATCCCATGGACCTTCAGAAAGTGGGATCCCATCATCATGACATCTAGCCT 337
QY 123 GGCAGTATCATCTGTGTTGTCCTCATCAAGGTGATTCGGATAAAATACTACTTCT 182

Db 338 GGCAGTATCATCTGTGTTGTCCTCATCAAGGTGATTCGGATAAAATACTACTTCT 397
QY 183 CTCGGGAGCCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGAGCTG 242
Db 398 CTCGGGAGCCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGAGCTG 457
QY 243 TCCTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCGAGTGGC 302
Db 458 TCCTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCGAGTGGC 517
QY 303 AGTCGCGCTCTCCAAAGAACCGATCCACACTGACAGTGTGTGGACTCGGCCACACAGGAATG 362
Db 518 AGTCGCGCTCTCCAAAGAACCGATCCACACTGACAGTGTGTGGACTCGGCCACACAGGAATG 577
QY 363 GTTCTCTGCTGTTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 422
Db 578 GTTCTCTGCTGTTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 637
QY 423 GGGCTACACAGCAAAACCCACTTTCAGAGCTGTGTGAGATTGGCCCGACAGCATCTGA 482
Db 638 GGGCTACAG-----CAGAGCTGTGAGATTGGCCCGACAGCATCTGA 682
QY 483 TGTGTTGAAATCACAGAAACAGCCAGAGCTTCGCATGCGGAACTCAAGTGGGCCCTG 542
Db 683 TGTGTTGAAATCACAGAAACAGCCAGAGCTTCGCATGCGGAACTCAAGTGGGCCCTG 742
QY 543 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTCCCTGTGGGAAGAGCTTGAAGACCC 602
Db 743 TCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTCCCTGTGGGAAGAGCTTGAAGACCC 802
QY 603 CGCTGTGTGGTGGGAGGAGGCTCTGTGTGATTCTTTGGCCTTGGCAGGTCAAGATCA 662
Db 803 CGCTGTGTGGTGGGAGGAGGCTCTGTGTGATTCTTTGGCCTTGGCAGGTCAAGATCA 862
QY 663 GTACGACAAACAGCAGCTGTGTGAGGAGAGGATCTGTGACACCCCACTGGGTCTCACGGC 722
Db 863 GTACGACAAACAGCAGCTGTGTGAGGAGAGGATCTGTGACACCCCACTGGGTCTCACGGC 922
QY 723 AGCCCACTGCTTTCAAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGGAGGCTCAGA 782
Db 923 AGCCCACTGCTTTCAAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGGAGGCTCAGA 982
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATGTAATCAACCC 842
Db 983 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATGTAATCAACCC 1042
QY 843 CATGTACCCAAAGACAAATGACATCGCCCTCATGAAGCTGAGTTCACCACTCACTTTCTC 902
Db 1043 CATGTACCCAAAGACAAATGACATCGCCCTCATGAAGCTGAGTTCACCACTCACTTTCTC 1102
QY 903 AGGCACAGTCAGGCCCACTCTGTCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCCC 962
Db 1103 AGGCACAGTCAGGCCCACTCTGTCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCCC 1162
QY 963 ACTCTGATCATTTGATGGGCTTTTGAAGAGATGAGGAGGAGTGTCTGACATACT 1022
Db 1163 ACTCTGATCATTTGATGGGCTTTTGAAGAGATGAGGAGGAGTGTCTGACATACT 1222
QY 1023 GCTCAGGCGTCAGTCCAGGTCATTGACAGCACACAGCGTGAATGACAGCATGGTACCA 1082
Db 1223 GCTCAGGCGTCAGTCCAGGTCATTGACAGCACACAGCGTGAATGACAGCATGGTACCA 1282
QY 1083 GGGGAAAGTCACCGAAGAGATGATGTGTGAGGAGTCCCGAAGGGGGTGTGACACCTG 1142
Db 1283 GGGGAAAGTCACCGAAGAGATGATGTGTGAGGAGTCCCGAAGGGGGTGTGACACCTG 1342
QY 1143 CCAGGTCAGAGTGGTGGGCGCTGATGATGACAGTGTGATGAGTGGGTCAT 1202
Db 1343 CCAGGTCAGAGTGGTGGGCGCTGATGATGACAGTGTGATGAGTGGGTCAT 1402
QY 1203 CGTTAGCTGGGCTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTC 1262

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 96.8%; Score 1263; DB 6; Length 2063;
Best Local Similarity 98.8%; Pred. No. 8.4e-292;
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 3 GGATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAAACCCCTGCGCAAAACCCCG 62
Db 218 GGATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAAACCCCTGCGCAAAACCCCG 277
QY 63 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGACCT 122
Db 278 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGACCT 337
QY 123 GCGAGTATCATATTGTGGTGTCTCATCAAGGTGATTCCTGATAAATACTACTTCCT 182
Db 338 GCGAGTATCATATTGTGGTGTCTCATCAAGGTGATTCCTGATAAATACTACTTCCT 397
QY 183 CTGGGGAGAGCTCTCCATCTTCCCGAGGAGAGCTGTGTGACGGAGAGCTGGAGCTG 242
Db 398 CTGGGGAGAGCTCTCCATCTTCCCGAGGAGAGCTGTGTGACGGAGAGCTGGAGCTG 457
QY 243 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCGCTCAGTGGC 302
Db 458 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCGCTCAGTGGC 517
QY 303 AGTCGGCTCTCCAGGAGCCGATCCACACTGAGGTGTGGACTCGGCCACAGGGAACTG 362
Db 518 AGTCGGCTCTCCAGGAGCCGATCCACACTGAGGTGTGGACTCGGCCACAGGGAACTG 577
QY 363 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
Db 578 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 637
QY 423 GGGGTACAGAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 482
Db 638 GGGGTACAG-----CAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 682
QY 483 TGTGTGAAATCAGAAACAGCAGGAGCTTCGATGCGGAACTCAAGTGGGCCCTG 542
Db 683 TGTGTGAAATCAGAAACAGCAGGAGCTTCGATGCGGAACTCAAGTGGGCCCTG 742
QY 543 TCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTTGTGGAGAGGCTGGAAGACCCC 602
Db 743 TCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTTGTGGAGAGGCTGGAAGACCCC 802
QY 603 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATCTTGTGGCTTGGCAGGTACAGATCCA 662
Db 803 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATCTTGTGGCTTGGCAGGTACAGATCCA 862
QY 663 GTACGACAAACAGACGCTCTGTGAGGAGAGCTCTGGACCCCGACCTGGGTCTCAGCGC 722
Db 863 GTACGACAAACAGACGCTCTGTGAGGAGAGCTCTGGACCCCGACCTGGGTCTCAGCGC 922
QY 723 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGGGGCAGGCTCAGA 782
Db 923 AGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGGGGCAGGCTCAGA 982
QY 783 CAAGTGGGAGCTTCCCATCTCGCTGGCTGTGGCCAAAGATCATCATTAATTCACACC 842
Db 983 CAAGTGGGAGCTTCCCATCTCGCTGGCTGTGGCCAAAGATCATCATTAATTCACACC 1042
QY 843 CATGTACCCCAAGACAAATGATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 902
Db 1043 CATGTACCCCAAGACAAATGATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 1102
QY 903 AGGCACAGTCAAGGCCATCTCTGTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 962
Db 1103 AGGCACAGTCAAGGCCATCTCTGTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1162

Search completed: June 8, 2004, 01:15:54
Job time : 5333 secs

QY 963 ACTCTGGATCATTGGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1022
Db 1163 ACTCTGGATCATTGGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1222
QY 1023 GCTGCAGGGCTCAGTCCAGGTCAATTGACAGCACACCGGTGCAATGACAGACGATGCGTACCA 1082
Db 1223 GCTGCAGGGCTCAGTCCAGGTCAATTGACAGCACACCGGTGCAATGACAGACGATGCGTACCA 1282
QY 1083 GGGGGAAGTACCGAGAAAGATGATGTGCAGGCATCCCGGAAGGGGGTGTGACACCTG 1142
Db 1283 GGGGGAAGTACCGAGAAAGATGATGTGCAGGCATCCCGGAAGGGGGTGTGACACCTG 1342
QY 1143 CCAGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGCGCATGTGTGGGCAT 1202
Db 1343 CCAGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGCGCATGTGTGGGCAT 1402
QY 1203 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTC 1262
Db 1403 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTC 1462
QY 1263 AGCCTATCTCAACTGGATCTACAAATGCTTGGGAAGGCTGAGCTG 1305
Db 1463 AGCCTATCTCAACTGGATCTACAAATGCTTGGGAAGGCTGAGCTG 1505

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47; Search time 23 Seconds
(without alignments)
976.404 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNIWYWKAEI 435

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID
1	2338	99.8	435	US-09-008-271A-6
2	2337	99.8	437	US-09-851-588-8
3	2196	93.8	423	US-03-656-002-2
4	2188	93.4	406	US-09-851-588-6
5	677.5	28.9	492	US-09-685-166A-895
6	676.5	28.9	492	US-09-342-749-2
7	676.5	28.9	492	US-09-691-840-2
8	655.5	28.0	454	US-09-518-046-2
9	588.5	25.1	417	US-09-820-002-4
10	584	24.9	455	US-03-261-416-2
11	580	24.8	376	US-09-820-002-2
12	574	24.5	416	US-09-000-846-2
13	571	24.4	798	1 US-02-200-900A-2
14	571	24.4	798	1 PCT-US94-00616-2
15	568	24.3	418	1 US-08-508-448C-25
16	568	24.3	418	4 US-09-370-838-82
17	568	24.3	418	4 US-09-370-838-83
18	566	24.2	418	4 US-09-370-838-62
19	558.5	23.8	283	3 US-08-807-151-1
20	558.5	23.7	283	1 US-09-478-957-1
21	531.5	22.7	232	1 US-08-508-448C-19
22	512.5	21.9	256	2 US-09-027-337-3
23	512.5	21.9	256	4 US-03-644-600-3
24	512.5	21.9	256	4 US-09-654-600A-3
25	511	21.8	638	2 US-08-681-151-3
26	507.5	21.7	255	3 US-08-944-483-67
27	503.5	21.5	285	4 US-09-023-942A-26

28	486	20.8	248	3	US-08-944-483-63	Sequence 63, Appl
29	483	20.6	314	4	US-09-023-942A-6	Sequence 6, Appl
30	482.5	20.6	284	4	US-09-387-375-7	Sequence 7, Appl
31	478	20.4	314	3	US-09-008-271A-3	Sequence 3, Appl
32	478	20.4	314	4	US-09-907-794A-257	Sequence 257, App
33	478	20.4	314	4	US-09-905-125A-257	Sequence 257, App
34	478	20.4	314	4	US-09-902-775A-357	Sequence 257, App
35	477	20.4	312	4	US-09-023-942A-4	Sequence 4, Appl
36	475	20.3	407	4	US-09-734-675-4	Sequence 4, Appl
37	472.5	20.2	317	4	US-09-386-629-7	Sequence 7, Appl
38	472.5	20.2	317	4	US-09-907-794A-263	Sequence 263, App
39	472.5	20.2	317	4	US-09-905-125A-263	Sequence 263, App
40	472.5	20.2	317	4	US-09-902-775A-263	Sequence 263, App
41	470	20.1	902	4	US-09-644-600-10	Sequence 10, Appl
42	470	20.1	902	4	US-09-654-600A-10	Sequence 10, Appl
43	467.5	20.0	290	4	US-09-386-653A-7	Sequence 7, Appl
44	466.5	19.9	405	4	US-09-734-675-2	Sequence 2, Appl
45	465.5	19.9	250	3	US-08-944-483-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Moran-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLN0113
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 99.8%; Score 2338; DB 3; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-234; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 1;

QY 1 MDPDSQPLNSLDVPLKRPIMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 60
DB 1 MDPDSQPLNSLDVPLKRPIMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 60

QY 61 LCGQPLHPIPRKQDCGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 120
DB 61 LCGQPLHPIPRKQDCGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 120

QY 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
DB 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240
DB 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240

QY 241 AAHCFRKHDTVFNWKRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300
DB 241 AAHCFRKHDTVFNWKRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300

QY 301 SGTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360

QY 361 QGEVTEKMCAGIPGGVDTCQDGGPLMYQSDQVHVGIVSWGCGGPGSTFGVTKV 420
DB 361 QGEVTEKMCAGIPGGVDTCQDGGPLMYQSDQVHVGIVSWGCGGPGSTFGVTKV 420

QY 421 SAYLNWIYNWKAEL 435
DB 421 SAYLNWIYNWKAEL 435

RESULT 2
US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/542,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/556,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-8

Query Match 99.8%; Score 2337; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.5e-234; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 0;

QY 2 DPDSQPLNSLDVPLKRPIMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 61
DB 4 DPDSQPLNSLDVPLKRPIMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 63

QY 62 CQOPLHFIPRKQDCGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 121
DB 62 CQOPLHFIPRKQDCGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 123

DB 64 CQOPLHFIPRKQDCGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 123

QY 122 FSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 181

DB 124 FSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 183

QY 182 LGSLSVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241

DB 184 LGSLSVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 243

QY 242 AHCFRKHDTVFNWKRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 301

DB 244 AHCFRKHDTVFNWKRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLOFPLTF 303

QY 302 GTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 361

DB 304 GTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 363

QY 362 GEVTEKMCAGIPGGVDTCQDGGPLMYQSDQVHVGIVSWGCGGPGSTFGVTKV 421

DB 364 GEVTEKMCAGIPGGVDTCQDGGPLMYQSDQVHVGIVSWGCGGPGSTFGVTKV 423

QY 422 AYLNIWYINWKAEL 435

DB 424 AYLNIWYINWKAEL 437

RESULT 3
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 93.8%; Score 2196; DB 4; Length 423;
Best Local Similarity 97.1%; Pred. No. 6.9e-220;
Matches 408; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 16 PLRKPIMPETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFLCGOPLHFIPRKQLC 75

DB 4 PCANPVSPPRPSPESVGIPITIIALLSLASIIIVVLLIKVILDKYFLCGOPLHFIPRKQLC 63

QY 76 DGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGNWFSACFDNFTALAE 135

DB 64 DGLDCEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGNWFSACFDNFTALAE 123

QY 136 TACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPCLSGSLVSLHCLACG 195

DB 124 TACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPCLSGSLVSLHCLACG 183

QY 196 KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNW 255

DB 184 KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNW 243

QY 256 VRAGSDKLGSPFSLAVAKIIIEFNPMYKDNIALMKLOPLTFSGTVRPICLPFFDEE 315
Db 244 VRAGSDKLGSPFSLAVAKIIIEFNPMYKDNIALMKLOPLTFSGTVRPICLPFFDEE 303
QY 316 LTPATPLMIIGWGTQKQNGKXSDILLOASQVIDSTRCNADDAQGEVTEKMKCAGIPE 375
Db 304 LTPATPLMIIGWGTQKQNGKXSDILLOASQVIDSTRCNADDAQGEVTEKMKCAGIPE 363
QY 376 GGVDTCCGSDGGLMYQSDQWHVVGIVSGYCGCGPSTPGVYTKVSAYLNIYNNWKAE 435
Db 364 GGVDTCCGSDGGLMYQSDQWHVVGIVSGYCGCGPSTPGVYTKVSAYLNIYNNWKAE 423

RESULT 4

US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David C.
; APPLICANT: Gish, Kurt E.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-68829-1/DJB/JUD/RMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

Query Match 93.4%; Score 2188; DB 4; Length 406;
Best Local Similarity 100.0%; Pred. No. 4.4e-219;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 30 VGPIIIALLSLASIIIVVLIKVIDKYFLCGOPLHFIPRKQDCGELDCPLGEDEEH 89
Db 1 VGPIIIALLSLASIIIVVLIKVIDKYFLCGOPLHFIPRKQDCGELDCPLGEDEEH 60
QY 90 CVKSPPEGPAVAVRLSKDRSTLQVLDSTAGNWFACDFNTEALAEATACROMGYSSKPTF 149
Db 61 CVKSPPEGPAVAVRLSKDRSTLQVLDSTAGNWFACDFNTEALAEATACROMGYSSKPTF 120
QY 150 RAVEIGPQDLDVVEITENSQELMRNSSGFCLSGLVSLHCLACGKSLKTPRVVVGEEA 209
Db 121 RAVEIGPQDLDVVEITENSQELMRNSSGFCLSGLVSLHCLACGKSLKTPRVVVGEEA 180
QY 210 SVDSWPMQVSIQYDKQHVCGGSLDHPHVLTAHCFRKHDTDFVFNKVRAGSDKLGSPFSL 269
Db 181 SVDSWPMQVSIQYDKQHVCGGSLDHPHVLTAHCFRKHDTDFVFNKVRAGSDKLGSPFSL 240
QY 270 AVAKIIIEFNPMYKDNIALMKLOPLTFSGTVRPICLPFFDEELTPATPLMIIGWGF 329
Db 241 AVAKIIIEFNPMYKDNIALMKLOPLTFSGTVRPICLPFFDEELTPATPLMIIGWGF 300
QY 330 TKQNGKXSDILLOASQVIDSTRCNADDAQGEVTEKMKCAGIPEGGVDTCCGSDGGL 389
Db 301 TKQNGKXSDILLOASQVIDSTRCNADDAQGEVTEKMKCAGIPEGGVDTCCGSDGGL 360
QY 390 MYQSDQWHVVGIVSGYCGCGPSTPGVYTKVSAYLNIYNNWKAE 435
Db 361 MYQSDQWHVVGIVSGYCGCGPSTPGVYTKVSAYLNIYNNWKAE 406

RESULT 5

US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqul
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 28.9%; Score 677.5; DB 4; Length 492;
Best Local Similarity 39.4%; Pred. No. 1e-61;
Matches 151; Conservative 56; Mismatches 127; Indels 49; Gaps 13;
QY 75 CDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSTAGNWFACDFNTEALA 134
Db 133 CDGSHCPGEGDENRCVLIY--GP-----NFILOYSSQKSWHPVQDDWNNY 191
QY 135 ETACROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSG-----PCL 182
Db 182 RAACRDMGY--KNNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLVSLHCLACGKSL---KTPRVVCGEEASVDSEWQVSIQYDKQHVCGGSLDHPHVL 239
Db 233 SKAVSLRCLACGVNLNSRQSRIVGGSALPCAWPQVSLHVQNVHVCVGGSIITPEWIV 292
QY 240 TAAHCFRKH--TDVFNKVRAGSDKLGSPF---PSLAVAKIIIEFNPMY---PKNDIAL 291
Db 293 TAAHCFRKH--TDVFNKVRAGSDKLGSPF---PSLAVAKIIIEFNPMY---SHPNYDSKTKNDIAL 348
QY 292 MKLOPLTFSGTVRPICLPFFDEELTPATPLMIIGWGTQKQNGKXSDILLOASQVIDS 351
Db 349 MKLOPLTFNDLVKPYCLPFGNMMLOPEQLCWMISGNGATEEK--GKTSEVLNAAKVLIET 407
QY 352 TRCNADDAQGEVTEKMKCAGIPEGGVDTCCGSDGGLMYQSDQ--WHVVGIVSGYCGCG 410
Db 408 QRCNSRYVVDNLITPAMI CAGFLQGNVDSQCGDGGPLVTSNNNIWLLIGDTSWGSCK 467
QY 411 PSTPGVYTKVSAYLNIYNNWKAE 433
Db 468 AYRFGVYGVNVPFTDIYRQMK 490

RESULT 6
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.

```

; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match      28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.3e-61;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSFPEGPVAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALA 134
Db 133 CDGVSHCPGGEDENRCVRLY--GP-----NFIQVYSSQRKSNHPVQCDDWENYNG 181
QY 135 ETACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNMSG-----PCL 182
Db 182 RAACRDMGY--KNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKKLYHSDACS 232
QY 183 SGLSVSLHCLACGKSL---KTPRVVGGEEASVDSPWQVSIQYDKQHVCGGSIIDPHWVL 239
Db 233 SKAVVSLRACIACGVNLNRSRQRIVGGESALPGAWPQVSLHVQNVHVCGGSIITPEWIV 292
QY 240 TAAHCFERKH-TDVFNNKVRAGSDKLGSP---PSLAVAKIIIIIEFNPMY---PKNDIAL 291
Db 293 TAAHCFERKH-TDVFNNKVRAGSDKLGSP---PSLAVAKIIIIIEFNPMY---PKNDIAL 291
QY 292 MKLQPLFTFSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDS 351
Db 349 MKLQPLFTFNDLVKPVCLPNPGMWLQPEQLCWISGWGATEEK-GKTSEVLNAAKVLIIET 407
QY 352 TRCNADDAVQGEVTEKMKCAGIPEGVDTCQDGGPLMYQSDQ-WHVVGIVSVGYGCGG 410
Db 408 QRCNSRYVYDNLITPAMICAGFLQGNVDSQCGSGPLVTSKNNIWWLIGDTSWGGCGAK 467
QY 411 PSTPGVYTKVSAYLWNIYVNWKA 434
Db 468 AYRPGVYGVNVFTDIYRQMEAD 491

RESULT 8
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match      28.0%; Score 655.5; DB 3; Length 454;
Best Local Similarity 35.7%; Pred. No. 1.7e-59;
Matches 158; Conservative 73; Mismatches 149; Indels 63; Gaps 18;

QY 22 IPMETPRKVGIPFIALLSLAIIIVVLIKVLID---KY-----YFLCQPHFIPRKQL 74
Db 44 LPPEVFSQSSSLGIITALL-----ILALIGLHFDSCGKRCRSFKC---IELITR--- 92
QY 75 CDGELDCPLGEDEEHCVKSFPEGPVAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALA 134
Db 93 CDGVSDCKGGEDEYRC-----VRVGQNAVQLVFTAA--SWKTMCSDDWKGHYA 139
QY 135 ETACROMGYSS-----KPTFRAVEIGPDQDLVDVVEITENSQELMRNMSGPCL 182
```



```
Db 140 NVACQALGFPPSVSSDNLRVSSLEQGFREEFVSIIDHLLPDDKVTALHHSVYVREG---CA 196
Qy 183 SGLSVLSHLCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 241
Db 197 SGHVVTLOCTACGHRRGYSRRIVGGNMSLLSQWPQASLQFGYHLCGGSVITPLWITA 256
Qy 242 AHCFRKHDTVP---NWKVRAGSDKL--GSFPLAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLYLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
Qy 294 LQPLTFSTGVTRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGKVCWTSWGAT-EDGGDASPVLNHAAVPLISNKI 369
Qy 354 CNADDAVQGEVTERKMCAGIPEGVDTCQSGSGPLMYQ---SQMWHVGVISWGYGC 408
Db 370 CNHRDVYGGIISPSMLCAGYLTGGVDSQCGSGGGLVQERRLWKLVGATSGIGCAEVN 429
Qy 413 TPGVTVKVSAYLNWLYNWK 435
Db 430 KPGVTVTRVTSFLDWIHEQMERDL 452

RESULT 9
US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weinlu
; APPLICANT: Di, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO01194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-4

Query Match 25.1%; Score 588.5; DB 4; Length 417;
Best Local Similarity 31.5%; Pred. No. 1.4e-52;
Matches 140; Conservative 60; Mismatches 161; Indels 83; Gaps 11;

Qy 19 KPRIMETFRKVGIPITIIALLSLASIIIVVVLTKVILD---KY---YFLCGQPLHFIPRKQ 78
Db 15 REKVAALT---AGTLLLLTAIGAASWAIVALLR----- 45
Qy 79 LDCPLGEBEEHCVKSPFPGPAVALRSLKDRSTLQVLDATGNWFSACDNFTEALAE 138
Db 46 -----SDQE-----PLYPGVQSADARLMVFDKTEGTRLLCSRSRANAVAGLSC 90
Qy 139 ROMGVSSKPTFRAVEIGPDQDLDVVEITEN-----SQELRMENSSGPCLS 183
Db 91 EEMGF-----LEAL---THSELVDVTRTAGANGTSFGFCVDEGRLPHTQRLLEVISFSDCPR 142
Qy 184 GSLVLSHLCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 242
Db 143 GRFLAAIQDCQGRKLPVDRIYVGRDTSIGLRWPMQVSLRYDGAHLCCGSLLSGDWLTAA 202
Qy 243 HCFRKHDTVPN-WKVRAGSDKLGSFPLAVAKIIIEFNPMY-----KNDIALMK 293
Db 203 HCFPERNVLGRWRVFAQAQAQSPHGLQGVQAVVYHGGYLPFRDPNSENSENSDIALVH 262
Qy 294 LQPLTFSTGVTRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTR 353
Db 263 LSSPLPLETYIQVCLPRAAGQALVDGKICTVTGWGNT-QYVQQAQVGLQEARVPLISNDV 321
```

```
Qy 354 CNADDAVQGEVTERKMCAGIPEGVDTCQSGSGPLMYQ-----SQMWHVGVISWGYGC 408
Db 322 CNGADFYGNQIKPFMFCAGYPEGIDACQDQSGGPFVCEDSISRTPRWRLCGIVSWGTGC 381
Qy 409 GGFSTPGVTVKVSAYLNWLYNWK 432
Db 382 ALAQKPGVTVTKVSDFRWFQFAIK 405

RESULT 10
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: C'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Query Match 24.9%; Score 584; DB 3; Length 455;
Best Local Similarity 34.2%; Pred. No. 4.8e-52;
Matches 153; Conservative 73; Mismatches 150; Indels 72; Gaps 21;

Qy 22 IPMETFRKVGIPITIIALLSLASIIIVVVLTKVILD---KY---YFLCGQPLHFIPRKQL 74
Db 44 LPFEVFSQSSLGIIAL-----ILALAIGLIHFDCSGKYRCRSPK---IBLITR--- 92
Qy 75 CDGELDCPLGDEEHCVKSPFPGPAVALRSLKDRSTLQVLDATGNWFSACDNFTEALA 134
Db 93 CDGVSDCKDGEDVRC-----VRVGGQNAVQLQVFTAA--SWKTMCSDDWKGHYA 139
Qy 135 ETACROMGYSS-----KPTFRAVEIGPDQDLDVVEITENSQELRMENSSGPCPL 162
Db 140 NVACQALGFPPSVSSDNLRVSSLEQGFREEFVSIIDHLLPDDKVTALHHSVYVREG---CA 196
Qy 183 SGLVLSHLCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 241
Db 197 SGHVVTLOCTACGHRRGYSRRIVGGNMSLLSQWPQASLQFGYHLCGGSVITPLWITA 256
Qy 242 AHCFRKHDTVP---NWKVRAGSDKL--GSFPLAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLYLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
Qy 294 LQPLTFSTGVTRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGKVCWTSWGAT-EDGGDASPVLNHAAVPLIS--- 366
Qy 354 CNADDAVQGEVTERKMCAGIPEGVDTCQSGSGPLMYQSDQ--NHVGVISWGYG 407
Db 367 -NKDLPQGRVRRWHHLPLHALRGLPDGWRNWSQCGSGGGLVQERRLWKLVGATSGIG 425
Qy 408 CGGPSTPGVTVKVSAYLNWLYNWK 435
Db 426 CADVKNPGVTVTRVTSFLDWIHEQMERDL 453
```

```
RESULT 11
US-09-820-002-2
; Sequence 2, Application US/09820002
```

```
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-820-002-2

Query Match          24.8%; Score 580; DB 4; Length 376;
Best Local Similarity 30.8%; Pred. No. 9.5e-52;
Matches 132; Conservative 57; Mismatches 146; Indels 94; Gaps 9;

QY 19 KPRIMETKRGVPIPIIIALLSLAIIIVVLIKVILDKYFLCGQLHFIPIRKLCDGE 78
DB 15 RKVVAALT---AGTLLLLTAIGAASWAIIVALLR-----45

QY 79 LDCPLGEDEHCKVSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTALAEATAC 138
DB 46 -----SDQE-----PLYPVQVSSADARLMTVFKTEGTWRLLCSSRNARVAGLSC 90

QY 139 RONGYSSKPTTFAVEIGPQDLDVVEITENSQELMRNSSGCLSGSLVSLHCLACG-KS 197
DB 91 EENGFLS-----DCPRGRFLAACQDCGRK 116

QY 198 LKTPRVYGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAHCFRKHDTVDN-WKV 256
DB 117 LPVDRIVGGRDTSLGRWPQVSLRYDGAHLCGSLLSGDWVLTAAHCFPERNRVLSRWV 176

QY 257 RAGSDKLGSFPFLAVAKIIIEFNMPY-----KNDIALMKLPQLFSGTVRVC 308
DB 177 FAGAVASPHGLQLGVQVAVVHGGVLPFRDPSNSENNDIALVHLSGLPLTEYIQPVC 236

QY 309 LPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASQVIDSTRCNADDAYQGVETKM 368
DB 237 LPAAGQALVDGKICTVTGWGNT-QYYGQAGVLQEARVPIISNDVCGADFYGNQIKPKM 295

QY 369 MCAGIEGGVDTCCGDSGGPLMYQ-----SDQHVHVGVISWGYGCGGSTPGVTIKVSAY 423
DB 296 FCAGYEGGIDACQDSGGFFVCEDSISRTPRWRLCGIVSWGTGALAAKPGVYTKVSDF 355

QY 424 LNWYINVMK 432
DB 356 REWIFQAIK 364

RESULT 12
US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US

; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-820-002-2

Query Match          24.8%; Score 574; DB 2; Length 416;
Best Local Similarity 30.8%; Pred. No. 4.6e-51;
Matches 135; Conservative 62; Mismatches 151; Indels 90; Gaps 10;

QY 30 VGPIIIALLSLAIIIVVLIKVILDKYFLCGQLHFIPIRKLCDGELDPLGDEEH 89
DB 22 VGLLFTGTGAASWAIIVILLO-----SDQE- 48

QY 90 CVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDNFTALAEATACRQMY----- 143
DB 49 -----PLYQVQLSPGDSRLAVFDKTEGTWRLLCSSRNARVAGLGEEMGFLALAH 100

QY 144 -----SSKPTTFAVEIG-----PDQDLVDVVEITENSQELMRNSSGCLSGSLVSL 189
DB 101 SELDVRTAGANGTSGFCVDEGGLRLAQLLDVIVCD-----CPRGRFLTA 147

QY 190 HCLACG-KSLKTPRVYGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAHCFRKH 248
DB 148 TCQDCGRKRLPVDRIVGGQDSLSGRWPQVSLRYDGTDLHLCGSLLSGDWVLTAAHCFPER 207

QY 249 TDYFN-WKVPSAGSDKLGSFPFLAVAKIIIEFNMPY-----KNDIALMKLPQLT 299
DB 208 NRVLRSRVRVAGAVARTSPHAYQLGVQAVIYHGGVLPFRDPTIDENSNDIALVHLSLSP 267

QY 300 FSGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASQVIDSTRCNADDA 359
DB 268 LTEYIQVPCIPPAAGQALVDGKICTVTGWGNT-QYYGQAGVLQEARVPIISNEVCNSPDF 326

QY 360 YQGEVTEKMKCAGIEGGVDTCCGDSGGPLMYQ-----SDQHVHVGVISWGYGCGGSTP 414
DB 327 YGNQIKPKMF CAGYEGGIDACQDSGGFFVCEDSISGTSRWRRLCGIVSWGTGALARKP 386

QY 415 GVTYKVSAYLNWYINVMK 432
DB 387 GVTYKVTDFREWIFKAIK 404

RESULT 13
US-08-200-300A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
;; STREET: 87 CambridgePark Drive
;; CITY: Cambridge
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02140

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/200,900A
;; FILING DATE: 23-FEB-1994

;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meinerdt, Maureen C.
;; REGISTRATION NUMBER: 31,544
;; REFERENCE/DOCKET NUMBER: GI 5201-FWC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 876-1170 X8574
;; TELEFAX: (617) 876-5851

;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 798 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; PCT-US94-00616-2

Query Match 24.4%; Score 571; DB 1; Length 798;

Best Local Similarity 35.0%; Pred. No. 2.6e-50;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

```
QY 69 IPRKQLCDGELDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 IPLVNLCDGPHCKDGSDEAHCVRLP-NGTDSGLVQFR--IQSI-----WHVACAEN 485

QY 129 FTEALAEATACRQMGY-----SSKPTFRAVEIGPDQDLVVE-----ITENSQELMRNMSGP 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 WTTQISDDVCQLGLGTGNSVPTF-STGGGPVYNLTAPNGSLILTPSQO-----535

QY 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSPWQVSIQYDKQHVCGGSILD 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 CLEDSLILLQCNYSKCGKLVTCQVSPKIVGSDSREGAPWVVALYFDDQVCGASLVS 595

QY 235 PHWLTAHC-FRKHTDVFNWVRAGSKLGSF--PSLAVAKIIIIIEFNPMY----PKDND 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 RDWLVAACHVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRRKND 655

QY 289 IALMKLQPLTSGTVRPTCLPFDEELTPATPLIIGWGTQKNGKMSDILLQASVQV 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 656 IAMHLEKVNITYDIQIPLPEENVFPFGRCISAGWALIYQ-GSTADVLQEAADVPL 714

QY 349 IDSTRCNADDAYQGEVTEKMCAGIPEGVDTCQDGGSGPLMYQ-SDQWHVVGIVSWGYG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 LSNEKQ-QQMPENYNTENMVCAGYEAGVDSCQDGGSGPLMCCQENNRWLLAGVTSFGYQ 773

QY 408 CGGPSTPGVYTKVSAYLNWI 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 CALPNRPGVYARVPRFTEWI 793
```

RESULT 14

PCT-US94-00616-2

; Sequence 2, Application PC/TUS9400616

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

; NUMBER OF SEQUENCES: 33

; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/00616
;; FILING DATE:

;; CLASSIFICATION:
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 798 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; PCT-US94-00616-2

Query Match 24.4%; Score 571; DB 5; Length 798;

Best Local Similarity 35.0%; Pred. No. 2.6e-50;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

```
QY 69 IPRKQLCDGELDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 IPLVNLCDGPHCKDGSDEAHCVRLP-NGTDSGLVQFR--IQSI-----WHVACAEN 485

QY 129 FTEALAEATACRQMGY-----SSKPTFRAVEIGPDQDLVVE-----ITENSQELMRNMSGP 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 WTTQISDDVCQLGLGTGNSVPTF-STGGGPVYNLTAPNGSLILTPSQO-----535

QY 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSPWQVSIQYDKQHVCGGSILD 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 CLEDSLILLQCNYSKCGKLVTCQVSPKIVGSDSREGAPWVVALYFDDQVCGASLVS 595

QY 235 PHWLTAHC-FRKHTDVFNWVRAGSKLGSF--PSLAVAKIIIIIEFNPMY----PKDND 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 RDWLVAACHVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRRKND 655

QY 289 IALMKLQPLTSGTVRPTCLPFDEELTPATPLIIGWGTQKNGKMSDILLQASVQV 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 656 IAMHLEKVNITYDIQIPLPEENVFPFGRCISAGWALIYQ-GSTADVLQEAADVPL 714

QY 349 IDSTRCNADDAYQGEVTEKMCAGIPEGVDTCQDGGSGPLMYQ-SDQWHVVGIVSWGYG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 LSNEKQ-QQMPENYNTENMVCAGYEAGVDSCQDGGSGPLMCCQENNRWLLAGVTSFGYQ 773

QY 408 CGGPSTPGVYTKVSAYLNWI 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 CALPNRPGVYARVPRFTEWI 793
```

RESULT 15

US-08-508-448C-25

; Sequence 25, Application US/08508448C

; Patent No. 5804410

; GENERAL INFORMATION:

; APPLICANT: Kazuyoshi YAMAOKA et al.

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/508,448C

```

; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-508-448C-25

Query Match 24.3%; Score 568; DB 1; Length 418;
Best Local Similarity 32.4%; Pred. No. 2e-50;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

Qy 33 PIIALLSLASIIIVVLIKVLIDKYFLCGQLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYVVCIVVAGVILAVTIALLV---YFLAFQKSYFYRSSFQLLNVEYNSQLNSPATQE 72
Qy 84 -----GEDEHCVKSPPEG-----PAVAVRLSKD---RSTLOVLDSATGNWFSACFD 127
Db 73 YRFLSRIEISLITKPKESNLRNQFRAHAKLRQDGSGRADVVMKFOFTNNNGASM- 131
Qy 128 NFEALAEATACROMGYSSKPTFAVEIGPDQDLVDVVEITENSQELMRN--SSGCLSGS 185
Db 132 ---KSRIESVLRQMLNNS---GNLEINP--STEITSLTDQAAANWLINECGAGPDLI-- 180
Qy 186 LVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAACHF 245
Db 181 -----TLSEORILGGTEAEGSWPQVSLRLNNAHCGGSLINNMMILTAACHF 229
Qy 246 RKHTDVFNKVRAGSKLGSFFSL--AVAKIIIEFNPMYPKNDIALMKLOFPLTFSGT 303
Db 230 RSNNSPRDWIATSGIST--TFPKLRVRVRNLIHNNYKSATHENDIALVRLSNSVTFTKD 287
Qy 304 VRPICLPFDEELTPATPLWILGWGFTKQNGKMSDILLQASVQVVIDSTRCNADDAAYQGE 363
Db 288 IHSVCLPAATQNIPTPGSTAYVTGWGAQYAGHTVPE--LRQGGVRIISNDVCNAPHSYNGA 346
Qy 364 VTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQ--MHVVGIVSWGYGCGGPTPGVYTKYS 421
Db 347 ILGMLCAGVPGQGVQDACQDSCGGLVQEDSRRLWFIVGIVSWGDCQGLPKPKGVYTRVT 406
Qy 422 AYLNI 427
Db 407 AYLDWI 412
```

Search completed: June 7, 2004, 08:42:39

Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 20:10:49 ; Search time 573 Seconds
(without alignments)
9675.216 Million cell updates/sec

Title: US-10-030-688-1
Perfect score: 1305
Sequence: 1 atggtatctgacagtgtatca.....atgtctggaagctgagctg 1305

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002as: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1305	100.0	1305	4	AAD02556 Human ser
2	1303.4	99.9	2038	2	AXX87154 Human pro
3	1303	99.8	1314	7	ABX76354 Lung canc
4	1303	99.8	1314	9	ADB80524 Ovarian c
5	1303	99.8	2081	6	AAI72976 CJA8 pref
6	1303	99.8	2307	6	ABX76529 cDNA enco
7	1303	99.8	2307	7	ACF12937 Human cer
8	1301.8	99.8	2121	6	AAI64284 Human ser
9	1298.2	99.5	2137	4	ADL13114 Human mem
10	1297.4	99.4	1479	4	ADL13114 Human mem
11	1294	99.2	2165	6	ABZ35527 Human gen
12	1292	99.0	2070	3	AAZ90471 Human gen
13	1292	99.0	2079	4	AAI13169 Human tra
14	1292	99.0	2079	6	AAI72975 CJA8 cDNA
15	1292	99.0	2079	7	ABX57763 cDNA enco
16	1292	99.0	2079	8	ACF35041 Human tra
17	1292	99.0	2079	8	ADE48026 DNA enco
18	1263	96.8	2063	3	AAZ37099 Human PRO
19	1263	96.8	2063	4	AAI54396 DNA enco
20	1263	96.8	2063	4	AAI54396 DNA enco
21	1263	96.8	2063	4	AAI54396 DNA enco
22	1263	96.8	2063	4	AAI54396 DNA enco
23	1263	96.8	2063	6	ABX74433 Human CDN
					ABX11090 cDNA enco

24	1263	96.8	2063	7	ABX78692 Human PRO
25	1263	96.8	2063	7	ACA75664 Novel hum
26	1263	96.8	2063	7	ACA71144 Human sec
27	1263	96.8	2063	7	ACC87672 Human sec
28	1263	96.8	2063	7	ACC87058 Human sec
29	1263	96.8	2063	7	ACD04231 Human sec
30	1263	96.8	2063	7	ACA69562 cDNA enco
31	1263	96.8	2063	7	ACA90407 Novel hum
32	1263	96.8	2063	7	ACC89514 Human sec
33	1263	96.8	2063	7	ACA98305 Novel hum
34	1263	96.8	2063	7	ACA93947 Human sec
35	1263	96.8	2063	7	ACD15340 Human sec
36	1263	96.8	2063	7	ACD08927 Human sec
37	1263	96.8	2063	7	ACC96847 Human sec
38	1263	96.8	2063	7	ACF15568 Human sec
39	1263	96.8	2063	7	ACA72935 Human PRO
40	1263	96.8	2063	7	ACD03107 Novel hum
41	1263	96.8	2063	7	ACD01922 Novel hum
42	1263	96.8	2063	7	ACA92114 Novel hum
43	1263	96.8	2063	7	ACA89539 cDNA enco
44	1263	96.8	2063	7	ACA73549 Human sec
45	1263	96.8	2063	7	ACA05864 Human sec

ALIGNMENTS

RESULT 1
AAD02556
ID AAD02556 standard; cDNA; 1305 BP.
XX
AC AAD02556;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human seripancrin cDNA.
XX
KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cardiovascular disease; Gene therapy;
KW vaccine; cytostatic; cerebroprotective; vulnary; osteopathic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1305
FT FT /*tag= a
FT FT /product= "Human seripancrin protein"
FT FT /note= "The coding region does not include stop codon"
FT FT /partial
FT
XX WO200104141-A2.
XX 18-JAN-2001.
XX 04-JUL-2000; 2000WO-EP006211.
XX 12-JUL-1999; 99EP-00113428.
XX (MERE) MERCK PATENT GMBH.
XX Suendermann B, Hofmann U, Matzku S, Wilbert O;
XX WPI; 2001-147177/15.
XX P-PSDB; AAY72558.
XX New extracellular serine protease Seripancrin, useful for treating
XX cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
XX diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound
XX healing.
XX Claim 5; Page 37-39; 45pp; English.

XX The present invention relates to seripancrin polynucleotides, and
 CC polypeptides encoded by them. Seripancrin are members of serine protease
 CC family. This protein contains a transmembrane domain, a low density
 CC lipoprotein (LDL) domain, protease domain and a scavenger receptor
 CC cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the
 CC specificity of seripancrin's intra and intermolecular interactions. The
 CC polynucleotides and polypeptides of the invention are useful for treating
 CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary
 CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
 CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
 CC diseases. Seripancrin genes are useful in chromosome localization
 CC studies, as tools for tissue expression studies and also in gene therapy.
 CC The polypeptides of the invention are used for identifying agonists and
 CC antagonists useful for treating conditions associated with seripancrin
 CC imbalance. These polypeptides are also useful as vaccines. The present
 CC sequence is a cDNA coding for seripancrin protein. The seripancrin gene
 CC is located on human chromosome 11q22-q23
 XX

SQ Sequence 1305 BP; 293 A; 368 C; 372 G; 272 T; 0 U; 0 Other;

Query Match 100.0%; Score 1305; DB 4; Length 1305;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTTCGATGTCAAAACCCCTCGCAAAACC 60
 DB 1 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTTCGATGTCAAAACCCCTCGCAAAACC 60

QY 61 CGTATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATATAGCACTACTGAGC 120
 DB 61 CGTATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATATAGCACTACTGAGC 120

QY 121 CTGGCGAGTATCATATCTGTGTCTCTCATCAAGGTGATTCTCGATAAATACTACTTC 180
 DB 121 CTGGCGAGTATCATATCTGTGTCTCTCATCAAGGTGATTCTCGATAAATACTACTTC 180

QY 181 CTCTGGGAGCCTCTCCACTTCATCCCGAGGAGCTGTGTGACGGAGAGCTGGAC 240
 DB 181 CTCTGGGAGCCTCTCCACTTCATCCCGAGGAGCTGTGTGACGGAGAGCTGGAC 240

QY 241 TGTCCCTCTGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 300
 DB 241 TGTCCCTCTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 300

QY 301 GCAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGTGTGACTCGGCCACAGGGAAC 360
 DB 301 GCAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGTGTGACTCGGCCACAGGGAAC 360

QY 361 TGGTCTCTGCTGTTTCGAACTTCACAGAGCTCTCGCTGACAGAGCTGTAGGCAG 420
 DB 361 TGGTCTCTGCTGTTTCGAACTTCACAGAGCTCTCGCTGACAGAGCTGTAGGCAG 420

QY 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480
 DB 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480

QY 481 GATGTTGTTGAATACAGAAACAGCAGGAGCTTCGCATCGGAACCTCAAGTGGGCC 540
 DB 481 GATGTTGTTGAATACAGAAACAGCAGGAGCTTCGCATCGGAACCTCAAGTGGGCC 540

QY 541 TGTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTGCTGGGAAGAGCTGAAGACC 600
 DB 541 TGTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTGCTGGGAAGAGCTGAAGACC 600

QY 601 CCCCCTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTTCAGCATC 660
 DB 601 CCCCCTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTTCAGCATC 660

QY 661 CAGTACGAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGTCTCTCAG 720
 DB 661 CAGTACGAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGTCTCTCAG 720

QY 721 GCAGCCCACTGCTTACAGAAACATACCCGATGTTTCAACTGGAAGTGGGGCAGGCTCA 780
 DB 721 GCAGCCCACTGCTTACAGAAACATACCCGATGTTTCAACTGGAAGTGGGGCAGGCTCA 780

QY 781 GACAAACTGGGAGCTTCCATCCCTGGCTGGTGGCCAGATCATCATATTGAATTCAC 840
 DB 781 GACAAACTGGGAGCTTCCATCCCTGGCTGGTGGCCAGATCATCATATTGAATTCAC 840

QY 841 CCCATGTACCCCAAGACAAATGACATCGGCCCTCATGAAGCTCAGTCCCACTCACTTTC 900
 DB 841 CCCATGTACCCCAAGACAAATGACATCGGCCCTCATGAAGCTCAGTCCCACTCACTTTC 900

QY 901 TCAGGCACAGTACAGGCCCACTCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 960
 DB 901 TCAGGCACAGTACAGGCCCACTCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 960

QY 961 CCACTCTGGATCAATGGATGGGCTTTACGAAGCAGATGGAGGAGATGTCTGACATA 1020
 DB 961 CCACTCTGGATCAATGGATGGGCTTTACGAAGCAGATGGAGGAGATGTCTGACATA 1020

QY 1021 CTGCTCAGGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATTCAGACAGCATGCTAC 1080
 DB 1021 CTGCTCAGGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATTCAGACAGCATGCTAC 1080

QY 1081 CAGGGGAAAGTACCCAGAGAGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1140
 DB 1081 CAGGGGAAAGTACCCAGAGAGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1140

QY 1141 TGCCAGGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACCACTGGCATGTGGTGGGC 1200
 DB 1141 TGCCAGGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACCACTGGCATGTGGTGGGC 1200

QY 1201 ATCCTTAGCTGGGCTATGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTTC 1260
 DB 1201 ATCCTTAGCTGGGCTATGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTTC 1260

QY 1261 TCAGCCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
 DB 1261 TCAGCCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305

RESULT 2
 AAX87154
 ID AAX87154 standard; cDNA; 2038 BP.
 XX AC AAX87154;
 XX XX
 DT 27-SEP-1999 (first entry)
 XX XX
 DE Human protease HUPW-6 cDNA.
 XX XX
 KW Serine protease; human; HUPW-6; cell proliferation; cancer;
 KW immune disorder; inflammation; therapy; ss.
 XX OS Homo sapiens.
 XX XX
 FH Key Location/Qualifiers
 FT CDS 200..1507
 FT /*tag= a
 XX XX
 FN WO9936550-A2.
 XX XX
 PD 22-JUL-1999.
 XX XX
 PF 12-JAN-1999; 99WO-US000655.
 XX XX
 PR 16-JAN-1998; 98US-00008271.
 XX XX
 PA (INCY-) INCYTE PHARM INC.
 XX XX
 PI Bandman O, Hillman JL, Yue H, Guegler KJ, Corley NC, Tang YT;
 XX Shah P;

DR WPI; 1999-430616/36.
XX P-PSDB; AAY06437..
XX Novel human protease molecules useful in the treatment of developmental
PT disorders and/or cancers.
XX Claim 8; Page 86-87; 90pp; English.
XX This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel human
CC protease. HUPM-6 cDNA was initially identified in Incyte Clone 1337018
CC from the colon cDNA library COLNN0713 using a computer search for amino
CC acid sequence alignments. The present sequence is a consensus sequence
CC derived from overlapping and/or extended nucleic acid consensus sequence
CC Clones 1271725 (TESTTUT02), 1337018, 586982 and 588598 (UTRNN001). A
CC fragment comprising nucleotides 900-949 of the present sequence can be
CC used for hybridisation. This sequence encompasses an active site residue.
CC Northern analysis shows expression of HUPM-6 in gastrointestinal, and male
CC an female reproductive cDNA libraries. Approximately 65% of these
CC libraries are associated with neoplastic disorders, and 22% with the
CC immune response. The invention provides 12 new human proteases, i.e. HUPM
CC -1 to -12 (see AAY06432-43), and the polynucleotides encoding them (see
CC AAX87149-60). Also provided are vectors, host cells and methods for
CC producing HUPM polypeptides, as well as agonists and antagonists of
CC Methods for treating or preventing cell proliferative disorders and
CC immune disorders using HUPM or HUPM antagonists are claimed
XX
SQ Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 U; 0 Other;
Query Match 99.9%; Score 1303.4; DB 2; Length 2038;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGATCTCGACAGTCAACCTCTGAACAGCTCGATGTCAAACCCCTGGCGAAACC 60
DB 200 ATGGATCTCGACAGTCAACCTCTGAACAGCTCGATGTCAAACCCCTGGCGAAACC 259
QY 61 CCGATCCCCATCGACAGCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CCGATCCCCATCGACAGCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 319
QY 121 CTGGCGAGTATCATATGTTGTTGTCCTCATCAAGGTGATTCGGAATAATCTACTTTC 180
DB 320 CTGGCGAGTATCATATGTTGTTGTCCTCATCAAGGTGATTCGGAATAATCTACTTTC 379
QY 181 CTTCTGGGCGACCTCTCCACTTCTCCCGAGGACAGCTGTGTGACGGAGCTGGAC 240
DB 380 CTTCTGGGCGACCTCTCCACTTCTCCCGAGGACAGCTGTGTGACGGAGCTGGAC 439
QY 241 TGTCCCTTTGGGGAGGACGAGGACGACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTG 300
DB 440 TGTCCCTTTGGGGAGGACGAGGACGACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTG 499
QY 301 GCAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGGTGCTGGAGTCGGCCACAGGGAAC 360
DB 500 GCAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGGTGCTGGAGTCGGCCACAGGGAAC 559
QY 361 TGGTCTCTGCTGCTGTTTCGACAACTTCAGAAAGCTTCGCTGACAGACGCTGTAGGCAG 420
DB 560 TGGTCTCTGCTGCTGTTTCGACAACTTCAGAAAGCTTCGCTGACAGACGCTGTAGGCAG 619
QY 421 ATGGGCTACAGCAGCAAAACCCCTTCAGAGCTGTGGAGATGTCGCGCCAGACAGGATCTG 480
DB 620 ATGGGCTACAGCAGCAAAACCCCTTCAGAGCTGTGGAGATGTCGCGCCAGACAGGATCTG 679
QY 481 GATGTTGTTGAATACAGAAACAGCAGGAGCTTCGATCGCGAATCTCAAGTGGGCCC 540
DB 680 GATGTTGTTGAATACAGAAACAGCAGGAGCTTCGATCGCGAATCTCAAGTGGGCCC 739
QY 541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAAGAGCTGAAGACC 600
DB 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAAGAGCTGAAGACC 799
QY 601 CCCCCTGTGGTGGGAGGAGGAGCCCTCTGTGGATTCTTGGCCCTTGGCAGGTACGATC 660

DB 800 CCCCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTACGATC 859
QY 661 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCCTCAGC 720
DB 860 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCCTCAGC 919
QY 721 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTTGAAGGTGCGGGCAGGCTCA 780
DB 920 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTTGAAGGTGCGGGCAGGCTCA 979
QY 781 GACAACTGGGAGCTTCCCATCTGGCTGTGGCCAAAGATCATCATATGATTAATCAAC 840
DB 980 GACAACTGGGAGCTTCCCATCTGGCTGTGGCCAAAGATCATCATATGATTAATCAAC 1039
QY 841 CCCATGTACCCCAAGACAATGATCGCTCATGAAGTGCAGTTCCTCCACTCCTTTC 900
DB 1040 CCCATGTACCCCAAGACAATGATCGCTCATGAAGTGCAGTTCCTCCACTCCTTTC 1099
QY 901 TCAGGCACAGTCAAGGCCCATCTCTGTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACC 960
DB 1100 TCAGGCACAGTCAAGGCCCATCTCTGTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACC 1159
QY 961 CCCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATCGAGGAGAGATGTCTGACATA 1020
DB 1160 CCCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATCGAGGAGAGATGTCTGACATA 1219
QY 1021 CTGTCTGAGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATTCAGACGATGCGTAC 1080
DB 1220 CTGTCTGAGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATTCAGACGATGCGTAC 1279
QY 1081 CAGGGGAAAGTCCAGGAAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACC 1140
DB 1280 CAGGGGAAAGTCCAGGAAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACC 1339
QY 1141 TGCAGGCTGACAGTGTGGGCTTCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGC 1200
DB 1340 TGCAGGCTGACAGTGTGGGCTTCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGC 1399
QY 1201 ATCTTGTAGCTGGGCTATGCTGGGGGCTCCGAGACACCCAGGAGTATACCAAGGTC 1260
DB 1400 ATCTTGTAGCTGGGCTATGCTGGGGGCTCCGAGACACCCAGGAGTATACCAAGGTC 1459
QY 1261 TCAGCCCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1305
DB 1460 TCAGCCCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1504
RESULT 3
ABX76354
ID ABX76354 standard; DNA; 1314 BP.
XX AC ABX76354;
XX DT 02-APR-2003 (first entry)
XX DE Lung cancer-associated polynucleotide #218.
XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
XX KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX OS Unidentified.
XX FN WO200286443-A2.
XX PD 31-OCT-2002.
XX PF 18-APR-2002; 2002WO-US012476.
XX PR 18-APR-2001; 2001US-0284770P.

10-MAY-2001; 2001US-0290492P.
 09-NOV-2001; 2001US-0339245P.
 13-NOV-2001; 2001US-0350666P.
 29-NOV-2001; 2001US-0334370P.
 12-APR-2002; 2002US-0372246P.
 (BOSB-) EOS BIOTECHNOLOGY INC.
 Aziz N, Murray R;
 WPI; 2003-093161/08.
 P-PSDB; ABU56625.
 Detecting a lung cancer-associated transcript in a cell from a patient
 for treating lung cancer, by contacting a biological sample from the
 patient with a polynucleotide that exhibits increased or decreased
 expression in lung cancer.
 Claim 22; Page 353; 453pp; English.
 The invention relates to a method for detecting a lung cancer-associated
 transcript in a cell from a patient, comprising contacting a biological
 sample from the patient with a polynucleotide that selectively hybridizes
 to a sequence that is at least 80 % identical to a gene that exhibits
 increased or decreased expression in lung cancer samples. Lung cancer-
 associated polynucleotides and polypeptides are used for identifying a
 compound that modulates a lung cancer-associated polypeptide, for
 inhibiting proliferation of a lung cancer-associated cell to treat lung
 cancer in a patient and for treating a mammal having lung cancer by
 administering a modulatory compound identified. The methods are useful
 for treating lung cancer, such as small cell lung cancer, non-small cell
 lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 bronchiectasis. The genes, polynucleotides and polypeptides are useful
 for diagnostic purposes and as targets for screening for therapeutic
 compounds that modulate lung cancer, such as antibodies. Sequences
 ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
 invention
 Sequence 1314 BP; 297 A; 369 C; 373 G; 275 T; 0 U; 0 Other;
 Query Match 99.8%; Score 1303; DB 7; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 3 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTCGCAAAACCCCG 62
 9 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTCGCAAAACCCCG 68
 63 TATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATATAGCACTACTAGGCT 122
 69 TATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATATAGCACTACTAGGCT 128
 123 GGGAGTATCATATTTGGTGTCTCTCATCAGGTGATTCGGATATATCTACTTCT 182
 129 GGGAGTATCATATTTGGTGTCTCTCATCAGGTGATTCGGATATATCTACTTCT 188
 183 CTCGGGCGACCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGCTGGACTG 242
 189 CTCGGGCGACCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGCTGGACTG 248
 243 TCCCTTGGGGAGGAGAGAGACATGTGTCAAGACTTCCCGAAGGCGCTGCAGTGC 302
 249 TCCCTTGGGGAGGAGAGAGACATGTGTCAAGACTTCCCGAAGGCGCTGCAGTGC 308
 303 AGTCCGCTCTCCAGGACCGATCCACATCGCAGGTGCTGGACTCGGCCACAGGGAATG 362
 309 AGTCCGCTCTCCAGGACCGATCCACATCGCAGGTGCTGGACTCGGCCACAGGGAATG 368
 363 GTTCTCTGCTGTTTTCGACAACTTCACAGAGCTTCGCTGAGACAGGCTGTAGGCAGT 422
 369 GTTCTCTGCTGTTTTCGACAACTTCACAGAGCTTCGCTGAGACAGGCTGTAGGCAGT 428

QY 423 GGGTACAGCAGCAAAACCCCACTTTACAGCTGTGGAGATTGGCCACAGACCGAGTCTGGA 482
 DB 429 GGGTACAGCAGCAAAACCCCACTTTACAGCTGTGGAGATTGGCCACAGACCGAGTCTGGA 488
 QY 483 TGTGTTTGAATCACAGAAAACAGCAGGAGCTTCGCATCGCGAACTCAAGTGGGCGCTG 542
 DB 489 TGTGTTTGAATCACAGAAAACAGCAGGAGCTTCGCATCGCGAACTCAAGTGGGCGCTG 548
 QY 543 TCTCTAGGCTCCCTGTCTCTCCCTGCATCTCTTCTGCTGTGGAGAGCCTGAGACCCC 602
 DB 549 TCTCTAGGCTCCCTGTCTCTCCCTGCATCTCTTCTGCTGTGGAGAGCCTGAGACCCC 608
 QY 603 CCGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTTCAGCATCCA 662
 DB 609 CCGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTTCAGCATCCA 668
 QY 663 GTACGACAAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCACTGGGTCTCTCAGGC 722
 DB 669 GTACGACAAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCACTGGGTCTCTCAGGC 728
 QY 723 AGCCACTGCTTCAGGAAACATACCGATGTTCACCTGGAAGTGGCGCAGGCTCAGA 782
 DB 729 AGCCACTGCTTCAGGAAACATACCGATGTTCACCTGGAAGTGGCGCAGGCTCAGA 788
 QY 783 CAAACTGGGAGCTTCCCACTCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACCC 842
 DB 789 CAAACTGGGAGCTTCCCACTCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACCC 848
 QY 843 CATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGCAGTTCCTCCTCCTTCTC 902
 DB 849 CATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGCAGTTCCTCCTCCTTCTC 908
 QY 903 AGGCACAGTCAAGGCCATCTGTCTGCCCTCTTGTGATGAGGAGCTCACTCCAGCCACCC 962
 DB 909 AGGCACAGTCAAGGCCATCTGTCTGCCCTCTTGTGATGAGGAGCTCACTCCAGCCACCC 968
 QY 963 ACTCTGATCATTTGGATGGGGCTTTACGAAGCAGATGAGAGGGAAGATGTCTGACATCT 1022
 DB 969 ACTCTGATCATTTGGATGGGGCTTTACGAAGCAGATGAGAGGGAAGATGTCTGACATCT 1028
 QY 1023 GCTCAGGCGCTCAGTCCAGGCTATTGACAGCACACGCTGCAATGACAGCATGCTACCA 1082
 DB 1029 GCTCAGGCGCTCAGTCCAGGCTATTGACAGCACACGCTGCAATGACAGCATGCTACCA 1088
 QY 1083 GGGGAAAGTCAACGAGAAAGATGTGTGAGGATCCCGAAGGGGTGTGACACCTG 1142
 DB 1089 GGGGAAAGTCAACGAGAAAGATGTGTGAGGATCCCGAAGGGGTGTGACACCTG 1148
 QY 1143 CCAGGTGACAGTGGTGGGCGCTGTGATGATCAATCTGACCACTGATGCTGGTGGGCAT 1202
 DB 1149 CCAGGTGACAGTGGTGGGCGCTGTGATGATCAATCTGACCACTGATGCTGGTGGGCAT 1208
 QY 1203 CGTTAGCTGGGGCTATGCTCGGGGGCGCCAGCACACCCAGGAGTATACACCAAGTCTC 1262
 DB 1209 CGTTAGCTGGGGCTATGCTCGGGGGCGCCAGCACACCCAGGAGTATACACCAAGTCTC 1268
 QY 1263 AGCCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1305
 DB 1269 AGCCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1311

RESULT 4

ADB80524

ID ADB80524 standard; DNA; 1314 BP.

XX

AC ADB80524;

XX

DT 04-DEC-2003 (first entry)

XX

DE Ovarian cancer-associated transcript #45.

XX

KW cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;

Db 1003 AGCCCCACTCTTCAGAAACATACGATGTGTCACTGGAAGGTGCGGAGGCTCAGA 1062
 QY 783 CAAACTGGGCGAGCTTCCCATCCTCGCTGTGCGCAAGATCATCATCAATTGAATCAACCC 842
 Db 1063 CAAACTGGGCGAGCTTCCCATCCTCGCTGTGCGCAAGATCATCATCAATTGAATCAACCC 1122
 QY 843 CATGTATCCCAAGCAATGATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 902
 Db 1123 CATGTATCCCAAGCAATGATCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTC 1182
 QY 903 AGGCACAGTCAGCCCATCTGTCTGCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 962
 Db 1183 AGGCACAGTCAGCCCATCTGTCTGCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 1242
 QY 963 ACTGTGATCATTTGGATGGGCTTTACGAAGCAGATGAGGAGGAGATGTCTGACATCT 1022
 Db 1243 ACTGTGATCATTTGGATGGGCTTTACGAAGCAGATGAGGAGGAGATGTCTGACATCT 1302
 QY 1023 GCTGAGGCTGAGTCCAGTCAATGACACACAGCTGCAATGACAGGATGCGTACCA 1082
 Db 1303 GCTGAGGCTGAGTCCAGTCAATGACACACAGCTGCAATGACAGGATGCGTACCA 1362
 QY 1083 GGGGGAAGTCACCGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGACACCTG 1142
 Db 1363 GGGGGAAGTCACCGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGACACCTG 1422
 QY 1143 CAGAGTGACAGTGTGGGCTGCTGATGTACCAATCTGACAGTGGCAT 1202
 Db 1423 CAGAGTGACAGTGTGGGCTGCTGATGTACCAATCTGACAGTGGCAT 1482
 QY 1203 CGTTAGCTGGGCTATGCTGCGGGGCGGAGCAGCAGCAGGATATACCAAGGTCTC 1262
 Db 1483 CGTTAGCTGGGCTATGCTGCGGGGCGGAGCAGCAGCAGGATATACCAAGGTCTC 1542
 QY 1263 AGCTTATCTCACTGATCTACATGTCTGGAAGGCTGAGCTG 1305
 Db 1543 AGCTTATCTCACTGATCTACATGTCTGGAAGGCTGAGCTG 1585

RESULT 7

ACFI2937
 ID ACFI2937 standard; cDNA; 2307 BP.
 XX ACFI2937;
 XX
 AC AC
 XX
 DT 10-SEP-2003 (first entry)
 XX
 DE Human cervical cancer cell marker encoding cDNA SEQ ID NO:217.
 XX
 KW Human; cervical cancer; cervical cancer marker; cancer therapy;
 XX detection; gene therapy; vaccine; gene; ss.
 OS Homo sapiens.
 XX
 PN WO2002101075-A2.
 XX
 PD 19-DEC-2002.
 XX
 PF 12-JUN-2002; 2002WO-US018638.
 XX
 PR 13-JUN-2001; 2001US-0298155P.
 PR 13-JUN-2001; 2001US-0298159P.
 PR 14-NOV-2001; 2001US-0335936P.
 XX
 XX {MILL-} MILLENNIUM PHARM INC.
 XX PA
 XX Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
 PI Gannavarapu M, Glatk K, Heersch S;
 XX WPI; 2003-156967/15.
 DR P-PSDB; ABR92154.
 XX

PT New isolated nucleic acid molecule useful for detecting, characterizing,
 PT preventing and treating human cervical cancers, in various prognostic and
 PT diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
 XX
 PS Claim 1; Page 369; 386pp; English.
 XX

CC ACFI2828 to ACFI2947 encode the human cervical cancer marker proteins (I)
 CC given in ABR92047 to ABR92164. A higher level of expression of (I) than
 CC normal indicates the presence of cervical cancer. Also described: (1) a
 CC vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
 CC assessing (M1) whether a patient is afflicted with cervical cancer,
 CC comprising comparing the level of expression of a marker in a patient's
 CC sample, and the normal level of expression of the marker in a control non
 CC -cervical cancer sample, where a significant increase in the level of
 CC expression of the marker in the patient's sample relative to that in the
 CC control sample is an indication that the patient is afflicted with
 CC cervical cancer. (I) has cytostatic activity, and can be used in gene
 CC therapy and in vaccines. (I) is useful in detecting, characterizing,
 CC preventing and treating human cervical cancers. (I) may also be used in
 CC various prognostic and diagnostic assays, pharmacogenomics and in
 CC monitoring clinical trials

SQ Sequence 2307 BP; 555 A; 647 C; 614 G; 460 T; 0 U; 31 Other;

Query Match 99.8%; Score 1303; DB 7; Length 2307;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAACCCCTGGCAAAACCCG 62
 Db 283 GGATCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAACCCCTGGCAAAACCCG 342
 QY 63 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCCTACTGAGCCT 122
 Db 343 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCCTACTGAGCCT 402
 QY 123 GGGGAGTATCATATGTGTGTCTCTCATCAGGTGATCTGGATAAATACTACTTCT 182
 Db 403 GGGGAGTATCATATGTGTGTCTCTCATCAGGTGATCTGGATAAATACTACTTCT 462
 QY 183 CTGGGGCAGCCTCTCCATTTCCGAGAGAGAGCTGTGTGACGAGAGCTGAGCTG 242
 Db 463 CTGGGGCAGCCTCTCCATTTCCGAGAGAGAGCTGTGTGACGAGAGCTGAGCTG 522
 QY 243 TCCCTTGGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGGC 302
 Db 523 TCCCTTGGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGGC 582
 QY 303 AGTCCGCTCTCCAAGGACCGATCCACACTGCAAGTGTGGACTCGGCCACAGGAACTG 362
 Db 583 AGTCCGCTCTCCAAGGACCGATCCACACTGCAAGTGTGGACTCGGCCACAGGAACTG 642
 QY 363 GTTCTCTGCTGTTCGACAACTTCAAGAGCTCTCGCTGACAGCCTGTAGGAGAT 422
 Db 643 GTTCTCTGCTGTTCGACAACTTCAAGAGCTCTCGCTGACAGCCTGTAGGAGAT 702
 QY 423 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACCATCTGGA 482
 Db 703 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACCATCTGGA 762
 QY 483 TGTGTGTTGAAATCACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCTG 542
 Db 763 TGTGTGTTGAAATCACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCTG 822
 QY 543 TCTCTCAGGCTCCCTGCTCTCCCTGACATGTCTTGCTGTGGAAAGAGCTGAGACCCC 602
 Db 823 TCTCTCAGGCTCCCTGCTCTCCCTGACATGTCTTGCTGTGGAAAGAGCTGAGACCCC 882
 QY 603 CCGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGCCCTTGGCAGTCTGAGCATCA 662
 Db 883 CCGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGCCCTTGGCAGTCTGAGCATCA 942
 QY 663 GTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCTGCTCTCAGGC 722

Db 943 GTACACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGGTCTCTACGGC 1002
Qy 723 AGCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 782
Db 1003 AGCCCACTGCTTCAGGAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 1062
Qy 783 CAACTGGGAGCTCCCATCCCTGGCTGTGGCCAGATCATCATGATGAATTCACCC 842
Db 1063 CAACTGGGAGCTTCCTCCCTGGCTGTGGCCAGATCATCATGATGAATTCACCC 1122
Qy 843 CATGTACCCCAAGACAATGACATCCCTCATGAGCTGAGTTCCTCCCACTCACTTTCTC 902
Db 1123 CATGTACCCCAAGACAATGACATCCCTCATGAGCTGAGTTCCTCCCACTCACTTTCTC 1182
Qy 903 AGGCACGTACGGCCCATCTGTCTGCCCTTTCTTGATGAGGAGCTCATCCAGCCACCC 962
Db 1183 AGGCACGTACGGCCCATCTGTCTGCCCTTTCTTGATGAGGAGCTCATCCAGCCACCC 1242
Qy 963 ACTCTGGATCATGTGATGGGCTTTACGAAGCAGATGAGGAGATGTCTGCATACT 1022
Db 1243 ACTCTGGATCATGTGATGGGCTTTACGAAGCAGATGAGGAGATGTCTGCATACT 1302
Qy 1023 GCTGAGGCGTCAGTCCAGGTCAATGACAGCACACGGTGCATGACAGATGCGTACCA 1082
Db 1303 GCTGAGGCGTCAGTCCAGGTCAATGACAGCACACGGTGCATGACAGATGCGTACCA 1362
Qy 1083 GGGGAAGTCAACGAGATGATGTGTGAGGATCCCGAGGAGGTGTGGACACTG 1142
Db 1363 GGGGAAGTCAACGAGATGATGTGTGAGGATCCCGAGGAGGTGTGGACACTG 1422
Qy 1143 CCAGGTCACAGTGTGGGCGCTCATGATACCAATCTGACAGTGGCATGTGTGGGAT 1202
Db 1423 CCAGGTCACAGTGTGGGCGCTCATGATACCAATCTGACAGTGGCATGTGTGGGAT 1482
Qy 1203 CTTAGTGGGCTATGCTGCGGGGCCAGGACCCAGGAGTATACCAAGTCTC 1262
Db 1483 CTTAGTGGGCTATGCTGCGGGGCCAGGACCCAGGAGTATACCAAGTCTC 1542
Qy 1263 AGCCTATCTCAACTGTATCTCAATGTCTGGAAGCTGAGCTG 1305
Db 1543 AGCCTATCTCAACTGTATCTCAATGTCTGGAAGCTGAGCTG 1585
RESULT 8
AAI64284
ID AAI64284 standard; cDNA; 2121 Bp.
XX AAI64284;
AC AAI64284;
XX 07-MAY-2002 (first entry)
XX Human serine protease D-G cDNA.
XX Serine protease; D-G; human; zymogen; enzyme; cytostatic;
XX antiinflammatory; dermatological; anticoagulation; cancer; skin disorder;
XX neuropathic pain; inflammatory disorder; coagulation diathesis;
XX thrombosis; laundry detergent; skin care; gene therapy; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 277..1584
CDS /*tag= a
FT /*product= "serine protease D-G"
FT /*transl_except= (pos: 1357..1359, aa: Gln)
FT polyA_signal 2100..2105
FT /*tag= b
FT /*note= "does not conform to consensus sequence"
XX
XX W0200202011-A1.
XX
XX 10-JAN-2002.
PD

XX 08-JUN-2001; 2001WO-US018568.
XX 30-JUN-2000; 2000US-00607745.
XX (ORTH) ORTHO-MCNEIL PHARM INC.
XX Darrow AL, Qi J, Andrade-Gordon P;
XX WPI; 2002-106601/14.
XX P-PSDB; AAG78577.
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin disorders,
XX or neuropathic pain.
XX Claim 4; Fig 1A; 81pp; English.
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, antiinflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,
XX skin disorders, neuropathic pain, inflammatory disorders, or coagulation
XX diathesis/thrombosis. The polynucleotide encoding the protease is useful
XX for identifying, detecting or isolating mutant forms of DNA molecules
XX encoding the protease. The protease is useful for identifying modulators
XX of the functional protease. The D-G protein can be used for formulation
XX of compositions for laundry detergents and skin care products. Protease D
XX -G gene therapy may be used to introduce protease D-G into the cells of
XX target organisms. As the D-G protein is derived from a human, it is less
XX likely to produce an allergic reaction in sensitive individuals when used
XX in formulations for laundry detergents and skin care products. The
XX current sequence represents the human serine protease D-G cDNA
XX Sequence 2121 Bp; 487 A; 614 C; 584 G; 436 T; 0 U; 0 Other;
Query Match 99.8%; Score 1301.8; DB 6; Length 2121;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGATCTCTGACAGTGTCAACCTCTGAACAGCTCGATCTCAACCCCTGGCAAAACCC 60
Db 277 ATGATCTCTGACAGTGTCAACCTCTGAACAGCTCGATCTCAACCCCTGGCAAAACCC 336
Qy 61 CGTATCCCAATGGAGACCTTCAGAAAGTGGGATCCCATCATCTAGCACTACTGAGC 120
Db 337 CGTATCCCAATGGAGACCTTCAGAAAGTGGGATCCCATCATCTAGCACTACTGAGC 396
Qy 121 CTGCGGAGTATCATTTGTTGTTCTCTCATCAAGTGTATCTTGGATAAATACTACTTC 180
Db 397 CTGCGGAGTATCATTTGTTGTTCTCTCATCAAGTGTATCTTGGATAAATACTACTTC 456
Qy 181 CTCTCGGGGAGGCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC 240
Db 457 CTCTCGGGGAGGCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC 516
Qy 241 TGTCCTTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGCGCTGAGTG 300
Db 517 TGTCCTTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGCGCTGAGTG 576
Qy 301 GCAGTCCGCTCTCTCAAGGACCGATCCCACTGCACTGCACTGTGTGACTCGGCCACAGGAAC 360
Db 577 GCAGTCCGCTCTCTCAAGGACCGATCCCACTGCACTGCACTGTGTGACTCGGCCACAGGAAC 636
Qy 361 TGGTTCTCTGCTGTCTTTCGACAACTTCACAGAGCTCTCGCTGAGAGAGCTGTAGCAG 420
Db 637 TGGTTCTCTGCTGTCTTTCGACAACTTCACAGAGCTCTCGCTGAGAGAGCTGTAGCAG 696

QY 421 ATGGGCTACAGCAGCAAAACCACTTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTG 480
 Db 697 ATGGGCTACAGCAGCAAAACCACTTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTG 756
 QY 481 GATGTTGTTGAATACAGAAACACAGCAGCAGGAGCTTCGATCGGNACTCAAGTGGGCC 540
 Db 757 GATGTTGTTGAATACAGAAACACAGCAGCAGGAGCTTCGATCGGNACTCAAGTGGGCC 816
 QY 541 TGCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTGAAGACC 600
 Db 817 TGCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGAAGAGCCTGAAGACC 876
 QY 601 CCCGTTGTTGGTGGGAGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGATC 660
 Db 877 CCCGTTGTTGGTGGGAGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGATC 936
 QY 661 CAGTAGCAAAACAGCAGCAGCTGTGTGGAGGAGCATCTCGACCCCACTGGGTCTCTCAGG 720
 Db 937 CAGTAGCAAAACAGCAGCAGCTGTGTGGAGGAGCATCTCGACCCCACTGGGTCTCTCAGG 996
 QY 721 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGCAGGCTCA 780
 Db 997 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGCGGCAGGCTCA 1056
 QY 781 GACAACTGGCAGCTTCCATCCCTGGCTGTGGCCAGATCATCATCAATCAAC 840
 Db 1057 GACAACTGGCAGCTTCCATCCCTGGCTGTGGCCAGATCATCATCAATCAAC 1116
 QY 841 CCCATGTACCCCAAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTCTTC 900
 Db 1117 CCCATGTACCCCAAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTCTTC 1176
 QY 901 TCAGGCAGATCAGGCGCATCTGTCTGCCCTTCTTGTAGGAGCTCATCCAGCCACC 960
 Db 1177 TCAGGCAGATCAGGCGCATCTGTCTGCCCTTCTTGTAGGAGCTCATCCAGCCACC 1236
 QY 961 CCACTCTGATCATTTGGATGGGCTTTACAGAGCAAGTGGAGGAAGATGCTGACATA 1020
 Db 1237 CCACTCTGATCATTTGGATGGGCTTTACAGAGCAAGTGGAGGAAGATGCTGACATA 1296
 QY 1021 CTGCTCAGCGGTGAGTCAAGTCAATGACAGACACCGGTGCAATGCGAGATCGGTAC 1080
 Db 1297 CTGCTCAGCGGTGAGTCAAGTCAATGACAGACACCGGTGCAATGCGAGATCGGTAC 1356
 QY 1081 CAGGGGGAAGTCAACGAGAGATGATGTGAGGACATCCCGAAGGGGTGTGACACC 1140
 Db 1357 CTGGGGGAAGTCAACGAGAGATGATGTGAGGACATCCCGAAGGGGTGTGACACC 1416
 QY 1141 TGCCAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGTGTGGGC 1200
 Db 1417 TGCCAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGTGTGGGC 1476
 QY 1201 ATCGTTAGCTGGGCTATGGCTGCGGGGCCCCGACGCCAGGAGTATACCAAGTTC 1260
 Db 1477 ATCGTTAGCTGGGCTATGGCTGCGGGGCCCCGACGCCAGGAGTATACCAAGTTC 1536
 QY 1261 TCAGGCTATCTCAACTGGAATCTACAAATGTCTGGAAGGCTGAGCTG 1305
 Db 1537 TCAGGCTATCTCAACTGGAATCTACAAATGTCTGGAAGGCTGAGCTG 1581

RESULT 9

AAD13114

ID AAD13114 standard; DNA; 2137 BP.

XX AC

XX AAD13114;

XX DT

XX 16-OCT-2001 (first entry)

XX DE

XX Human membrane-type serine protease (MTSP) 3 DNA.

XX KW

Human; transmembrane serine protease; membrane-type serine protease;

KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
 lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
 ds.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT CDS 261..1574
 FT /*tag= a
 FT /product= "Human transmembrane serine protease (MTSP) 3"
 PN W0200157194-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 02-FEB-2001; 2001WO-US003471.
 XX
 XX 03-FEB-2000; 2000US-0179982P.
 PR 18-FEB-2000; 2000US-0183542P.
 PR 22-JUN-2000; 2000US-0213124P.
 PR 26-JUL-2000; 2000US-0220970P.
 PR 08-SEP-2000; 2000US-00657986.
 PR 22-SEP-2000; 2000US-0234840P.
 XX
 PA (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;
 PI WPI; 2001-488877/53.
 DR P-PSDB; AAE06931.
 XX
 PT Novel single chain polypeptide comprising protease domain of type-II
 PT membrane-type serine protease or its catalytically active portion useful
 PT for treating and preventing cancer and tumor.
 XX
 PS Claim 14; Page 197-199; 256pp; English.
 XX

XX The invention relates to transmembrane serine proteases and their
 CC corresponding nucleotides and the protease domain of a type-II membrane-
 CC type serine protease (MTSP). MTSP is useful for identifying compounds
 CC that modulate or inhibit its proteolytic activity and for formulating a
 CC medicament for treating neoplastic disease. MTSP and its corresponding
 CC nucleotides are useful in preventing or treating tumours or cancers such
 CC as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
 CC diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
 CC marker for tumour development, growth and/or progression and as
 CC immunogens to generate antibodies that specifically bind to it. MTSP DNA
 CC is useful in a yeast two-hybrid system and in gene therapy. The present
 CC sequence is a DNA encoding human MTSP3 protein
 XX

SQ Sequence 2137 BP; 508 A; 612 C; 589 G; 428 T; 0 U; 0 Other;

Query Match 99.5%; Score 1298.2; DB 4; Length 2137;
 Best Local Similarity 99.8%; Pred.No. 0;
 Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 GGATCTCGACAGTCAACCTCTGAACAGCTTCGATGTCAAAACCCCTCGCAAAACCCCG 62
 Db 269 GGATCTCGACAGTCAACCTCTGAACAGCTTCGATGTCAAAACCCCTCGCAAAACCCCG 328
 QY 63 TATCCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
 Db 329 TATCCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 388
 QY 123 GCGAGTATCATATTGTGTTCTCTCATCAAGGTGATTCTCGATAAATACTACTTCT 182
 Db 389 GCGAGTATCATATTGTGTTCTCTCATCAAGGTGATTCTCGATAAATACTACTTCT 448
 QY 183 CTGCGGGAGCCTTCTCACTTATCCGAGGAGAGCTGTGTGACGAGAGCTGACTG 242
 Db 449 CTGCGGGAGCCTTCTCACTTATCCGAGGAGAGCTGTGTGACGAGAGCTGACTG 508
 QY 243 TCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGC 302

61	CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATGCACTACTGAGC	120
121	CTGGGAGTATCATCATTTGTGTGTCTCATCAAGTGATTTCTGGATAAATACTACTTC	180
121	CTGGCAGTATCATCATTTGTGTGTCTCATCAAGTGATTTCTGGATAAATACTACTTC	180
181	CTTCGGGGCAGCTCTCCATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC	240
181	CTTCGGGGCAGCTCTCCATTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGGAC	240
241	TGTCCTTTGGGGAGGACGAGACACATGTGTCAAGAGCTTCCCAGAAAGGCTCGACGT	300
241	TGTCCTTTGGGGAGGACGAGAGACATGTGTCAAGAGCTTCCCAGAAAGGCTCGACGT	300
301	GCAGTCGCGCTTCACAGGACCGATCCACACTGCAGGTGTGAGATCGGCCACAGGGAAC	360
301	GCAGTCGCGCTTCACAGGACCGATCCACACTGCAGGTGTGAGATCGGCCACAGGGAAC	360
361	TGGTCTCTGCTGTTCGCAAACTTCACAGAGCTCTCGCTGACAGACCTGTAGGCAG	420
361	TGGTCTCTGCTGTTCGCAAACTTCACAGAGCTCTCGCTGACAGACCTGTAGGCAG	420
421	ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGCACAGATCTG	480
421	ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGCACAGATCTG	480
481	GATGTTGTTGAAATCACHAAAAAGCCAGAGACTTCGCATGCGGAACTCAAGTGGGCC	540
481	GATGTTGTTGAAATCACHAAAAAGCCAGAGACTTCGCATGCGGAACTCAAGTGGGCC	540
541	TGCTCTCAGGCTCCCTGCTCTCCCTGCTCACTGTCTTGCCTGTGGGAAGAGCTGAAGAC	600
541	TGCTCTCAGGCTCCCTGCTCTCCCTGCTCACTGTCTTGCCTGTGGGAAGAGCTGAAGAC	600
601	CCCCGTGTGTTGGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGTCAAGATC	660
601	CCCCGTGTGTTGGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGTCAAGATC	660
661	CAGTACGACAAACAGCAGCTGTGGAGGAGAGCATCTCGAGCCCACTGGGTCTCAAG	720
661	CAGTACGACAAACAGCAGCTGTGGAGGAGAGCATCTCGAGCCCACTGGGTCTCAAG	720
721	GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCA	780
721	GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCA	780
781	GACAACTGGGAGCTTCCATCCCTGGCTGTGCGCAAGATCATCATTTGAATTCAC	840
781	GACAACTGGGAGCTTCCATCCCTGGCTGTGCGCAAGATCATCATTTGAATTCAC	840
841	CCCATGTACCCAAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCCACTCACTTC	900
841	CCCATGTACCCAAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCCACTCACTTC	900
901	TCAGGACAGTCAGGCCCATCTGTCTGCCCTTTTGTAGAGAGCTCACTCCAGCCACC	960
901	TCAGGACAGTCAGGCCCATCTGTCTGCCCTTTTGTAGAGAGCTCACTCCAGCCACC	960
961	CCACTCTGGATCATTTGATGGGCTTTACGACAGCAATGGAGGGAAGATCTGTGACATA	1020
961	CCACTCTGGATCATTTGATGGGCTTTACGACAGCAATGGAGGGAAGATCTGTGACATA	1020
1021	CTGTGTCAGGCGTCAGTCCAGGTCATTTGACAGCACCGGTGCAATGACAGAGTCGCTAC	1080
1021	CTGTGTCAGGCGTCAGTCCAGGTCATTTGACAGCACCGGTGCAATGACAGAGTCGCTAC	1080
1081	CAGGGGAGTCCAGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGAACC	1140
1081	CAGGGGAGTCCAGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGAACC	1140
1141	TGCCAGGTCACAGTGTGGGCCCTTGATGTACCAATCTGACACGTGGCATGTGGTGGC	1200

Db	1141	TGCCAGGTGACAGTGGTGGGCCCTCATGTACCAATCTGACCAGTGGCATGTGGTGGC	1200
Qy	1201	ATCGTTAGCTGGGGCTATGGCTCGGGGGCCCCAGACCCCCAGAGGATATACACCAAGTC	1260
Db	1201	ATCGTTAGCTGGGGCTATGGCTCGGGGGCCCCAGACCCCCAGAGGATATACACCAAGTC	1260
Qy	1261	TCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGCT	1299
Db	1261	TCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGAT	1299

RESULT 11

ABZ35523	
ID	ABZ35523 standard; cDNA; 2165 BP.
XX	
XX	ABZ35523;
XX	
XX	05-FEB-2003 (first entry)
XX	
DE	Human gene expression profile polynucleotide
XX	
XX	Human; artery; endothelium; umbilical;
KW	bronchial epithelium; prostate; muscle;
KW	tumour; microarray; genome mapping; anti-
KW	gene expression; gene; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200274979-A2.
XX	
PD	26-SEP-2002.
XX	
PF	20-MAR-2002; 2002WO-US008456.
XX	
XX	20-MAR-2001; 2001US-0276947P.
PR	
XX	(ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
PA	
XX	Wan J, Wang Y;
PI	
XX	WPI; 2002-740862/80.
DR	
XX	
PT	New gene expression profile generated f
PT	epithelial, and muscle cell types, usin
PT	pathologies involving alterations of ge
XX	
PS	Disclosure: Page 759-760; 850pp; English

The invention relates to a gene expression profile comprising one or more genes (AB234899-AB235692) and generated from a cell type. The cell type is a coronary artery endothelium, umbilical artery or vein endothelium, aortic endothelium, dermal microvascular endothelium, pulmonary artery endothelium, myometrium microvascular endothelium, keratinocyte epithelium, bronchial epithelium, mammary epithelium, prostate epithelium, renal cortical epithelium, renal proximal tubule epithelium, small airway epithelium, renal epithelium, umbilical artery smooth muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle, dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes, aortic smooth muscle, mesangial cells, coronary artery smooth muscle, bronchial smooth muscle, uterine smooth muscle, lung fibroblast, osteoblasts or prostate stromal cell. The gene expression profile is used for determining the level of RNA expression for a sample, determining the phenotype of a cell and distinguishing cell types. The gene or a protein expression profile is useful in identifying disease pathologies involving alterations of gene expression. The assessment of expression profiles may provide meaningful information with respect to tumour type and stage, treatment methods, and prognosis. The gene or protein expression profile may also be used for creating microarrays. The microarray is useful for genetic and physical mapping of genomes, DNA sequencing, genetic or medical diagnosis, genotyping of organisms, confirming cell or tissue identifications and in identifying promising antibiotics, antiviral or antifungal agents

1020	ACTGCTGACGGCGTCCAGTCCAGGTCATTGACAGACACAGCGTGCATATCAGACAGATGCGTA	1079
1330	ACTGCTGACGGCGTCCAGTCCAGGTCATTGACAGACACAGCGTGCATATCAGACAGATGCGTA	1389
1080	CCAGGGGGAAGTCAACCGAGAAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACAC	1139
1390	CCAGGGGGAAGTCAACCGAGAAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACAC	1449
1140	CTGCCAGGCTGACAGTGTGGGCCCCCTGATGTACAAATCTGACAGTGGCATGTGTGGG	1199
1450	CTGCCAGGCTGACAGTGTGGGCCCCCTGATGTACAAATCTGACAGTGGCATGTGTGGG	1509
1200	CATCGTTAGCTGGGCTATGCTGCGGGGGCCCGAGCACCCCGAGAGTATACACCAAGGT	1259
1510	CATCGTTAGCTGGGCTATGCTGCGGGGGCCCGAGCACCCCGAGAGTATACACCAAGGT	1569
1260	CTCAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGTG	1305
1570	CTCAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGTG	1615
RESULT 32		
AAZ90471		
ID	AAZ90471	standard; cDNA; 2070 BP.
XX	AAZ90471;	
AC	AAZ90471;	
XX	AAZ90471;	
DT	06-JUN-2000	(first entry)
XX		
DE		Cancer specific gene (CSG) sequence (clone ID 1283171).
XX		
XX		CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
KW		endometrial; uterine; lung; cytotoxic; ss.
XX		
OS		Homo sapiens.
XX		
PN	WO200012758-A1.	
XX	09-MAR-2000.	
PD		
XX		
PF	01-SEP-1999;	99WO-US019655.
XX		
PR	02-SEP-1998;	98US-0098880P.
XX		
PA	(DIAD-)	DIADEXUS LLC.
XX		
PI	Salceda S, Sun Y, Recipon H, Caferkey R;	
XX		
DR	WPI; 2000-256657/22.	
XX		
PT		Diagnosing, staging, monitoring, imaging and treating cancer especially
PT		gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT		involves measuring cancer specific gene levels in cells and body fluids.
XX		
PS	Claim 9; Page 45-46; 58pp; English.	
XX		
CC		The invention relates to detecting, diagnosing metastasis and staging
CC		cancer by measuring levels of cancer specific genes (CSG) in cells,
CC		tissues or body fluids. Their remission and progression, decreases and
CC		increases in CSG levels, is also monitored, by periodic sample analysis.
CC		The methods are useful for detecting cancers, especially gynecologic
CC		cancers which include ovarian, breast, endometrial and uterine cancer and
CC		lung cancer. Antibodies against the CSGs labeled with paramagnetic ions
CC		or a radioisotope is useful for imaging cancer and when conjugated with a
CC		cytotoxic agent are useful for treating cancer. The present sequence
CC		represents a CSG sequence (clone ID: 1283171 and Gene ID: 332459)
XX		
SQ	Sequence 2070 BP; 473 A; 599 C; 577 G; 421 T; 0 U; 0 Other;	
Query Match 99.0%; Score 1292; DB 3; Length 2070;		
Best Local Similarity 99.9%; Pred. No. 0;		
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;		
1	ATGGATCTTGACAGTCAATCACTCTGACAGCCTCGATGTCAAAACCCCTCGGCAACCC	60
310	ATGGATCTTGACAGTCAATCACTCTGACAGCCTCGATGTCAAAACCCCTCGGCAACCC	369
61	CGTATCCCATGAGACCTTTCAGAAAGGTGGGGATCCCATCATCATAGCACTACTGAGC	120
370	CGTATCCCATGAGACCTTTCAGAAAGGTGGGGATCCCATCATCATAGCACTACTGAGC	429
121	CTGGCAGTATCATCATTTGTGTGCTCTCATCAAGGTGATCTGGATAAATACTACTTC	180
430	CTGGCAGTATCATCATTTGTGTGCTCTCATCAAGGTGATCTGGATAAATACTACTTC	489
181	CTCTGGGGCAGCTCTCCACTTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC	240
490	CTCTGGGGCAGCTCTCCACTTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC	549
241	TGTCCTTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCAAGTG	300
550	TGTCCTTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTGCAAGTG	609
301	GCAGTCGGCTCTCCAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAAC	360
610	GCAGTCGGCTCTCCAGGACCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAAC	669
361	TGTTCTCTGCTGTTTCGACAACTTTCAGAAAGCTCTGCTGTGACAGCCTGTAGGCAG	420
670	TGTTCTCTGCTGTTTCGACAACTTTCAGAAAGCTCTGCTGTGACAGCCTGTAGGCAG	729
421	ATGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGTTGGCCCGACAGCACTG	480
730	ATGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGTTGGCCCGACAGCACTG	789
481	GATGTTGTTGAATACAGAAACACCCAGGA-GTTTCGATCCGGAATCAAGTGGGCC	539
790	GATGTTGTTGAATACAGAAACACCCAGGAGCTTCGATCGGAACTCAAGTGGGCC	849
540	CTGTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGGAAGCCTGAAGAC	599
850	CTGTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGGAAGCCTGAAGAC	909
600	CCCCGTGTGGTGGGAGGAGCCCTGTGGATCTTGGCTTGGCAGTCAAGAT	659
910	CCCCGTGTGGTGGGAGGAGCCCTGTGGATCTTGGCTTGGCAGTCAAGAT	969
660	CCAGTACGAAACAGCAGTCTGTGGAGGAGCATCTTGGACCCCTCGGCTCTCAC	719
970	CCAGTACGAAACAGCAGTCTGTGGAGGAGCATCTTGGACCCCTCGGCTCTCAC	1029
720	GGCAGCCACTGCTCAGGAACATACCATGTTTCACTGGAAGGTGGCGGAGGCTC	779
1030	GGCAGCCACTGCTCAGGAACATACCATGTTTCACTGGAAGGTGGCGGAGGCTC	1089
780	AGACAACTGGGAGCTTCCCATCTGCTGCTGGCCAGATCATCATTAATTCAA	839
1090	AGACAACTGGGAGCTTCCCATCTGCTGCTGGCCAGATCATCATTAATTCAA	1149
840	CCCCATGTACCCAAAGACATGACATGCGCCTCATGAAGTGCAGTTCCTCACTTT	899
1150	CCCCATGTACCCAAAGACATGACATGCGCCTCATGAAGTGCAGTTCCTCACTTT	1209
900	CTCAGGCACAGTCAAGGCCCATCTGCTGCGCTCTTTGATGAGGAGCTCACTCCAGCCAC	959
1210	CTCAGGCACAGTCAAGGCCCATCTGCTGCGCTCTTTGATGAGGAGCTCACTCCAGCCAC	1269
960	CCCACTCTGATCATTTGATGGGCTTTACGAAGCAAGTATGAGGAGAGATGTTGACAT	1019
1270	CCCACTCTGATCATTTGATGGGCTTTACGAAGCAAGTATGAGGAGAGATGTTGACAT	1329

QY 3 GGATCTGACAGTGTCAACCTCTGAAAGCCTCGATGTCAAACCCCTGGCGAAACCCCG 62
DB 222 GATCTGACAGTGTCAACCTCTGAAAGCCTCGATGTCAAACCCCTGGCGAAACCCCG 281
QY 63 TATCCCATGAGAGCTTCAGAAAGTGGGATCCCATCATATAGCACTACTGAGCCT 122
DB 282 TATCCCATGAGAGCTTCAGAAAGTGGGATCCCATCATATAGCACTACTGAGCCT 341
QY 123 GCGAGTATCATCTGTTGTTCTCATCAAGGTGATTCGGATAAATACTACTTCT 182
DB 342 GCGAGTATCATCTGTTGTTCTCATCAAGGTGATTCGGATAAATACTACTTCT 401
QY 183 CTGCGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 242
DB 402 CTGCGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 461
QY 243 TCCCTTGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 302
DB 462 TCCCTTGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 521
QY 303 AGTCCGCTCTCCAGAGCCGATCCACACTGCAAGTGTGGACTCGGCCACAGGGAACCTG 362
DB 522 AGTCCGCTCTCCAGAGCCGATCCACACTGCAAGTGTGGACTCGGCCACAGGGAACCTG 581
QY 363 GTTCTCTGCTGTTTGGCAAACTTCACAGAGCTCTCGCTGAGAGCAGCTGTAGGCAGAT 422
DB 582 GTTCTCTGCTGTTTGGCAAACTTCACAGAGCTCTCGCTGAGAGCAGCTGTAGGCAGAT 641
QY 423 GGGCTACAGCAGCAACCACTTTGAGAGCTGTGAGAGTGGCCAGACAGGATCTGGA 482
DB 642 GGGCTACAGCAGCAACCACTTTGAGAGCTGTGAGAGTGGCCAGACAGGATCTGGA 701
QY 483 TGTGTTGAAATCACAGAAAACAGCCAGAGCTTCGATGCGGAACCTCAAGTGGGCCCTG 542
DB 702 TGTGTTGAAATCACAGAAAACAGCCAGAGCTTCGATGCGGAACCTCAAGTGGGCCCTG 761
QY 543 TCTCTCAGCTCCCTGCTCTCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 762 TCTCTCAGCTCCCTGCTCTCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
QY 603 CCGTGTGTTGGTGGGAGGAGGCTCTGTGATTTCTGGCTTTGGCAGGTCAGCATCCA 662
DB 822 CCGTGTGTTGGTGGGAGGAGGCTCTGTGATTTCTGGCTTTGGCAGGTCAGCATCCA 881
QY 663 GTACAGCAACAGCAGCTGTGAGGAGGAGCATCTGCAACCCCTGCTGCTCCTAC--GG 721
DB 882 GTACAGCAACAGCAGCTGTGAGGAGGAGCATCTGCAACCCCTGCTGCTCCTACCGGG 941
QY 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGCGAGGCTCAG 781
DB 942 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGCGAGGCTCAG 1001
QY 782 ACAAACTGGGAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
DB 1002 ACAAACTGGGAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061
QY 842 CCATGTACCCCAAGACATGACATGCGCTTCAAGCTGAGTCCCATCTCACTTTCT 901
DB 1062 CCATGTACCCCAAGACATGACATGCGCTTCAAGCTGAGTCCCATCTCACTTTCT 1121
QY 902 CAGGCACAGTCAGGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
DB 1122 CAGGCACAGTCAGGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1181
QY 962 CACTCTGATCATTTGATGGGCTTTAGAGCAGATGGAGGAAGATGCTGACATAC 1021
DB 1182 CACTCTGATCATTTGATGGGCTTTAGAGCAGATGGAGGAAGATGCTGACATAC 1241
QY 1022 TGCTCAGGCGTCACTCCAGGTCATTGACAGCAGCAGCGTGAATGACAGCATCGTACC 1081
DB 1242 TGCTCAGGCGTCACTCCAGGTCATTGACAGCAGCAGCGTGAATGACAGCATCGTACC 1301

QY 1082 AGGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1141
DB 1302 AGGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1361
QY 1142 GCAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA 1201
DB 1362 GCAGGGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA 1421
QY 1202 TCGTTAGCTGGGCTATGCTGCGGGGCCCCGAGGACCCCGAGGATATACCAAGGTCT 1261
DB 1422 TCGTTAGCTGGGCTATGCTGCGGGGCCCCGAGGACCCCGAGGATATACCAAGGTCT 1481
QY 1262 CAGCTATCTCAACTGGAATCAATGTCTGGAAGCTGAGCTG 1305
DB 1482 CAGCTATCTCAACTGGAATCAATGTCTGGAAGCTGAGCTG 1525

RESULT 13
AAD13169
ID AAD13169 standard; DNA; 2079 BP.
XX
AC AAD13169;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human transmembrane protease serine 4 (TMPRSS4) DNA.
XX
KW Human; transmembrane serine protease; membrane-type serine protease;
KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW TMPRSS4; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 251..1522
FT /*tag= a
FT /product= "Human transmembrane protease, serine 4
FT (TMPRSS4)"
XX
FN WO200157194-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US003471.
XX
PR 03-FEB-2000; 2000US-0179982P.
PR 18-FEB-2000; 2000US-0183542P.
PR 22-JUN-2000; 2000US-0213124P.
PR 26-JUL-2000; 2000US-0220970P.
PR 08-SEP-2000; 2000US-00657986.
PR 22-SEP-2000; 2000US-0234840P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Madison EL, Ong EO, Yeh J;
XX
XX WPI; 2001-488877/53.
XX P-PSDB; AAE06944.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion useful
XX for treating and preventing cancer and tumor.
XX
XX Disclosure; Page 253-255; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II membrane-
XX type serine protease (MTSP). MTSP is useful for identifying compounds
XX that modulate or inhibit its proteolytic activity and for formulating a
XX medicament for treating neoplastic disease. MTSP and its corresponding
XX nucleotides are useful in preventing or treating tumours or cancers such
XX as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in

CC diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
CC marker for tumour development, growth and/or progression and as
CC immunogens to generate antibodies that specifically bind to it. MTSP DNA
CC is useful in a yeast two-hybrid system and in gene therapy. The present
CC sequence is a DNA encoding human transmembrane protease, serine 4
CC (TMPRSS4) protein
XX
SQ Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;
Query Match 99.0%; Score 1292; DB 4; Length 2079;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 1; Gaps 1;
QY 3 GGATCCTGACAGTGAATCAACTCTGAAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 62
Db 216 GGATCCTGACAGTGAATCAACTCTGAAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 275
QY 63 TATCCCATGAGACCTTCAGAAAG-GTGGGATCCCATCATCATAGCACTACTGAGCC 121
Db 276 TATCCCATGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY 122 TGGGAGTATCATATTGTGGTTCCTCATCAAGGTGATCTCGATAAATACTACTTCC 181
Db 336 TGGGAGTATCATATTGTGGTTCCTCATCAAGGTGATCTCGATAAATACTACTTCC 395
QY 182 TCTGGGGCAGCTCTCCACTTCATCCGAGAGCAGCTGTGTGCGGAGAGCTGGACT 241
Db 396 TCTGGGGCAGCTCTCCACTTCATCCGAGAGCAGCTGTGTGCGGAGAGCTGGACT 455
QY 242 GTCCCTTGGGGAGGAGCAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGG 301
Db 456 GTCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGG 515
QY 302 CAGTCCGCTCTCAAGAGCAGATCCACACTGAGGTGCTGGACTCGGCCACAGGAACT 361
Db 516 CAGTCCGCTCTCAAGAGCAGATCCACACTGAGGTGCTGGACTCGGCCACAGGAACT 575
QY 362 GGTCTCTGCTGTTTCAGCAACTTTCAGAGAGCTCTCGCTGAGACAGCTGTAGGAGA 421
Db 576 GGTCTCTGCTGTTTCAGCAACTTTCAGAGAGCTCTCGCTGAGACAGCTGTAGGAGA 635
QY 422 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTGG 481
Db 636 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTGG 695
QY 482 ATGTGTTGTAATCACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCCCT 541
Db 696 ATGTGTTGTAATCACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCCCT 755
QY 542 GTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTTCCTGTGGGAAGAGCCTGAAGACC 601
Db 756 GTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTTCCTGTGGGAAGAGCCTGAAGACC 815
QY 602 CCGGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTCAAGATCC 661
Db 816 CCGGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTCAAGATCC 875
QY 662 AGTAGCAAAACAGCAGCTGTGTGGAGGAGCATCTCGACCCCTCCTCGCTCAAGG 721
Db 876 AGTAGCAAAACAGCAGCTGTGTGGAGGAGCATCTCGACCCCTCCTCGCTCAAGG 935
QY 722 CAGCCCACTGTCTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGCGAGGCTCAG 781
Db 936 CAGCCCACTGTCTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGCGAGGCTCAG 995
QY 782 ACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCCAAGTATCATATCAATTCAACC 841
Db 996 ACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCCAAGTATCATATCAATTCAACC 1055
QY 842 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTTCT 901
Db 1056 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTTCT 1115

QY 902 CAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 961
Db 1116 CAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 1175
QY 962 CACTCTGATCATTTGGATGGGCTTTTACGAAGCAGATGGAGGAAGATGTCTGACATAC 1021
Db 1176 CACTCTGATCATTTGGATGGGCTTTTACGAAGCAGATGGAGGAAGATGTCTGACATAC 1235
QY 1022 TGTGAGGCGTCACTCCAGGTCAATTCAGACACACACCGTGCATTCAGACAGATGGTACC 1081
Db 1236 TGTGAGGCGTCACTCCAGGTCAATTCAGACACACACCGTGCATTCAGACAGATGGTACC 1295
QY 1082 AGGGGAAGTCAACCCAGAGAGATGTGTGAGGATCCCGAAGGGGTGTGGACACCT 1141
Db 1296 AGGGGAAGTCAACCCAGAGAGATGTGTGAGGATCCCGAAGGGGTGTGGACACCT 1355
QY 1142 GCCAGGCTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCCAGTGGCATGTGGGCA 1201
Db 1356 GCCAGGCTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCCAGTGGCATGTGGGCA 1415
QY 1202 TGTGAGTGGGCTATGCTGGGGGCCCCGAGCAGCAGCCAGAGTATACACCAAGGTCT 1261
Db 1416 TGTGAGTGGGCTATGCTGGGGGCCCCGAGCAGCAGCCAGAGTATACACCAAGGTCT 1475
QY 1262 CAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
Db 1476 CAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519
RESULT 14
AA172975
ID AAI72975 standard; cdNA; 2079 BP.
XX
AC AAI72975;
XX
DT 21-AUG-2002 (first entry)
XX
DE CUA8 cDNA.
XX
KW Gene: colorectal cancer; CGA7; CUA8; modulating protein; screening;
KW drug candidate; vaccine; ss.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
CD5 302..1522
FT /*tag= a
FT /product= "CUA8"
XX
XX US2002042067-A1.
XX
XX 11-APR-2002.
XX
XX 08-MAY-2001; 2001US-00851588.
XX
XX 17-AUG-2000; 2000US-00642252.
XX 06-SEP-2000; 2000US-00656002.
XX
XX (MACK/) MACK D.
XX (GISH/) GISH K C.
XX (WILS/) WILSON K E.
XX
XX Mack D, Gish KC, Willson KE;
XX
XX WPI; 2002-453647/48.
XX P-PSDB; AAG79358.
XX
XX Screening drug candidates for treating colorectal cancer, comprises
XX determining the effect of the candidate on the expression profile gene of
XX CGA7 or CUA8.
XX
XX Disclosure; Fig 5; 40pp; English.

CC The sequences given in AAI72973-76 encode the colorectal cancer proteins,
 CC CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer
 CC modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11
 CC (CJA8). These sequences may be used in the method of the invention for
 CC screening drug candidates. The method comprises adding a drug candidate
 CC to a cell that expresses an expression profile gene encoding CGA7, CJA8
 CC or fragments and determining the effect of the drug candidate on the
 CC expression of the expression profile gene. The new methods are used to
 CC screen bioactive agents for the ability to bind to or modulate the
 CC activity of CGA7 or CJA8 and evaluate the effect of a candidate
 CC colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit the
 CC activity of CGA7 or CJA8, respectively, and is used to screen for an
 CC agent that can interfere with the binding of CGA7 or CJA8 to the
 CC antibody. The antibody can be used to treat colorectal cancer. The
 CC antibody or a fragment of it is used to localize a therapeutic group to a
 CC colorectal cancer tissue, where the therapeutic group is a cytotoxic
 CC agent or a radioisotope. Antisense molecules are used to inhibit
 CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CJA8
 CC are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it are used
 CC to elicit an immune response. CGA7 or CJA8 is used to determine the
 CC prognosis of an individual with colorectal cancer. Nucleic acid encoding
 CC CGA7 or CJA8 can be used in vaccines
 XX

Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;

Query Match 99.0%; Score 1292; DB 6; Length 2079;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCCTGACGTATCAACCTCTTGAACAGCCTCGATGATCAAAACCCCTGCGCAACCCCG 62
 DB 216 GGATCCTGACGTATCAACCTCTTGAACAGCCTCGATGATCAAAACCCCTGCGCAACCCCG 275

QY 63 TATCCCATGGAGACTTCAGAAAG-GTGGGATCCCATCATCATCATCATCTAGCTAGGCC 121
 DB 276 TATCCCATGGAGACTTCAGAAAGGTGGGATCCCATCATCATCATCATCTAGCTAGGCC 335

QY 122 TGGCAGTATCATCATTTGGTGTGCTCATCAAGGTGATCTGGATAAATACTACTTCC 181
 DB 336 TGGCAGTATCATCATTTGGTGTGCTCATCAAGGTGATCTGGATAAATACTACTTCC 395

QY 182 TGTGGGGAGCCTCTCACTTATCCGAGAGCAGCTGTGTGACGAGAGCTGAGCT 241
 DB 396 TGTGGGGAGCCTCTCACTTATCCGAGAGCAGCTGTGTGACGAGAGCTGAGCT 455

QY 242 GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAGAGCTTCCCGAAGGGCTGCAGTGG 301
 DB 456 GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAGAGCTTCCCGAAGGGCTGCAGTGG 515

QY 302 CAGTCCGCTCTTCAAGAACCGATCCACACTGCAAGTGTCTGCACTCGGCCACAGGGAAT 361
 DB 516 CAGTCCGCTCTTCAAGAACCGATCCACACTGCAAGTGTCTGCACTCGGCCACAGGGAAT 575

QY 362 GGTCTCTGCTGTTTCCACAACTTCAGAAAGCTTCGCTGACACAGCCTGTAGGAGA 421
 DB 576 GGTCTCTGCTGTTTCCACAACTTCAGAAAGCTTCGCTGACACAGCCTGTAGGAGA 635

QY 422 TGGCTACAGCAGCAAAACCCACTTTTCAGAGTGTGGAGATTGGCCACAGCAGATCTGG 481
 DB 636 TGGCTACAGCAGCAAAACCCACTTTTCAGAGTGTGGAGATTGGCCACAGCAGATCTGG 695

QY 482 ATGTTGTTGAATTCACAAAACAGCAGGAGCTTCGATCGGAATCAAGTGGGCT 541
 DB 696 ATGTTGTTGAATTCACAAAACAGCAGGAGCTTCGATCGGAATCAAGTGGGCT 755

QY 542 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCCTGAAGACCC 601
 DB 756 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCCTGAAGACCC 815

QY 602 CCGTGTGCTGGTGGGAGAGGAGCCTCTGTGGATTCTTGGCCTTGGCAGTCAAGATCC 661
 DB 816 CCGTGTGCTGGTGGGAGAGGAGCCTCTGTGGATTCTTGGCCTTGGCAGTCAAGATCC 875

QY 662 AGTACGACAAACAGCAGCTGTGTGGAGGAGAGATCTCTGGACCCCTCTGGTCTCTCACGG 721
 DB 876 AGTACGACAAACAGCAGCTGTGTGGAGGAGAGATCTCTGGACCCCTCTGGTCTCTCACGG 935

QY 722 CAGCCCACTGCTTTCAGAAACATACCGATCTGTTCACCTGGAGGTGCGGGCAGGCTCAG 781
 DB 936 CAGCCCACTGCTTTCAGAAACATACCGATCTGTTCACCTGGAGGTGCGGGCAGGCTCAG 995

QY 782 ACAAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTAATCAACC 841
 DB 996 ACAAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTAATCAACC 1055

QY 842 CCAATGATCCCAAGACATGACATGCGCTCATGAAGCTGAGTCCCATCTCACTTTCT 901
 DB 1056 CCAATGATCCCAAGACATGACATGCGCTCATGAAGCTGAGTCCCATCTCACTTTCT 1115

QY 902 CAGGCACAGTCAGGCCCATCTGTCTCCCTTCTTTCATGAGGAGCTCACTCCAGCCACCC 961
 DB 1116 CAGGCACAGTCAGGCCCATCTGTCTCCCTTCTTTCATGAGGAGCTCACTCCAGCCACCC 1175

QY 962 CATCTGGATCATTTGATGGGCTTTTACGAAGCAGAAATGGAGGAAGATGTGTGACATAC 1021
 DB 1176 CATCTGGATCATTTGATGGGCTTTTACGAAGCAGAAATGGAGGAAGATGTGTGACATAC 1235

QY 1022 TGTGTGAGGCTCAGTCCAGGTCATTGACAGACACGCTGCAATGCAGACGATGGTACC 1081
 DB 1236 TGTGTGAGGCTCAGTCCAGGTCATTGACAGACACGCTGCAATGCAGACGATGGTACC 1295

QY 1082 AGGGGAAAGTCACCGAAGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACT 1141
 DB 1296 AGGGGAAAGTCACCGAAGAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACT 1355

QY 1142 GCCAGGGTCAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGATGTGGGGCA 1201
 DB 1356 GCCAGGGTCAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGATGTGGGGCA 1415

QY 1202 TCCTTAGCTGGGGCTATGGCTGCGGGGGCCCCGAGCACCCCGAGAGTATACACCAAGGTCT 1261
 DB 1416 TCCTTAGCTGGGGCTATGGCTGCGGGGGCCCCGAGCACCCCGAGAGTATACACCAAGGTCT 1475

QY 1262 CAGCCTATCTCACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1305
 DB 1476 CAGCCTATCTCACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1519

RESULT 15
 ABS57763
 ID ABS57763 standard; cDNA; 2079 BP.
 XX
 AC ABS57763;
 XX AC
 XX AC
 DT 05-FEB-2003 (first entry)
 XX
 DE cDNA encoding human colorectal cancer modulating protein CJA8.
 XX
 KW Colorectal cancer modulating protein; BCMP; CJA8; colorectal cancer;
 KW gene therapy; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PH Key
 FT CDS
 FT 251..1522
 FT /*tag= a
 FT /product= "CJA8"
 FT /note= "Colorectal cancer modulating protein"
 FT /transl_except= (pos:314..316, aa:Leu)
 XX
 PN US6455668-B1.
 XX
 PD 24-SEP-2002.
 XX
 PF 06-SEP-2000; 2000US-00656002.
 XX

PR 28-JAN-2000; 2000US-00493444.
PR 15-MAR-2000; 2000US-00525993.
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX Mack D, Gish KC, Wilson KE;
PI WPI; 2003-066245/06.
XX DR P-PSDB; ABG72428.
XX
XX Diagnosing colorectal cancer comprises determining the expression of a
PT gene encoding CUA8 in a first colon tissue of a first individual, and
PT comparing them to the expression of a gene encoding CUA8 in a second
PT normal tissue.
XX
XX Claim 11; Fig 1; 31pp; English.
XX
XX The invention describes a method of diagnosing colorectal cancer
CC comprising determining the expression of a gene encoding CUA8 or its
CC fragment in a first colon tissue of an individual, and comparing the
CC expression of the gene to the expression of the gene in normal tissue
CC from the individual or in tissue from a second unaffected individual. A
CC difference in the expression indicates that the first individual has
CC colorectal cancer. The CUA8 genes and proteins are useful for diagnosing
CC and prognosticating colorectal cancer. The methods are also useful for
CC screening candidate bioactive agents that can modulate colorectal cancer,
CC or for treating or inhibiting colorectal cancer in a patient e.g. by gene
CC therapy. This sequence encodes the human colorectal cancer modulating
CC protein CUA8
XX
XX Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;
XX
XX Query Match 99.0%; Score 1292; DB 7; Length 2079;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
XX
XX 3 GGATCCTCAGAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTCGCAAAACCCCG 62
XX
XX 216 GGATCCTCAGAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTCGCAAAACCCCG 275
XX
XX 63 TATCCCATGGAGACCTTCAGAAAG- GTGGGATCCCATCATCATAGCACTACTGAGCC 121
XX
XX 276 TATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
XX
XX 122 TGGCAGTATCATATTGGTTGTCTCATCAAGTGTATCTGTGATAAATACTACTTCC 181
XX
XX 336 TGGCAGTATCATATTGGTTGTCTCATCAAGTGTATCTGTGATAAATACTACTTCC 395
XX
XX 182 TCTGGGGCAGCCTCTCCACTTTCATCCGAGAGCAGCTGTGTGACGAGAGTGGACT 241
XX
XX 396 TCTGGGGCAGCCTCTCCACTTTCATCCGAGAGCAGCTGTGTGACGAGAGTGGACT 455
XX
XX 242 GTCCCTTGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTGG 301
XX
XX 456 GTCCCTTGGGGAGGACGAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTGG 515
XX
XX 302 CAGTCGGCCTCTCCAGGAGCCGATCCACTGAGGTGTGGACTTCGGCCACAGGGAAC 361
XX
XX 516 CAGTCGGCCTCTCCAGGAGCCGATCCACTGAGGTGTGGACTTCGGCCACAGGGAAC 575
XX
XX 362 GGTTCCTCGCCTGTTTTCGACAACTTCAAGAGTCTCTGCTGAGACAGCCTGTAGGAGA 421
XX
XX 576 GGTTCCTCGCCTGTTTTCGACAACTTCAAGAGTCTCTGCTGAGACAGCCTGTAGGAGA 635
XX
XX 422 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGTGTGGAGATTGGCCACAGGATCTGG 481
XX
XX 636 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGTGTGGAGATTGGCCACAGGATCTGG 695
XX
XX 482 ATGTTGTTGAAATCAAGAAAAACAGCCAGGAGCTTGGCATGGCACTCAAGTGGGCCCT 541
XX
XX 696 ATGTTGTTGAAATCAAGAAAAACAGCCAGGAGCTTGGCATGGCACTCAAGTGGGCCCT 755
XX
XX 542 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGTGGGAGAGCCCTGAAGACCC 601

Db 756 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGTGGGAGAGCCCTGAAGACCC 815
Qy 602 CCGTGTGTGGTGTGGGAGGAGGCTCTGTGTGATCTTTGGCTTTGGCAGGTGAGATCC 661
Db 816 CCGTGTGTGTGGTGTGGGAGGAGGCTCTGTGTGATCTTTGGCTTTGGCAGGTGAGATCC 875
Qy 662 AGTACGACAAACAGCAGCTCTGTGTGAGGAGAGCATCTGTGACCCCACTGGGTCTCCACGG 721
Db 876 AGTACGACAAACAGCAGCTCTGTGTGAGGAGAGCATCTGTGACCCCACTGGGTCTCCACGG 935
Qy 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTTGGGAGGTGCGGGAGGCTCAG 781
Db 936 CAGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTTGGGAGGTGCGGGAGGCTCAG 995
Qy 782 ACAAACCTGGSCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATGATTCATCAACC 841
Db 996 ACAAACCTGGSCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATGATTCATCAACC 1055
Qy 842 CCATGTACCCCAAGAGCAATGACATCCCTCATGAAGCTGCAGTTCCTCCACTCCTTTCT 901
Db 1056 CCATGTACCCCAAGAGCAATGACATCCCTCATGAAGCTGCAGTTCCTCCACTCCTTTCT 1115
Qy 902 CAGGCACTGACGGCCCATCTGTCTCCCTTTCTTGTATGAGGAGCTCATCCAGCCACCC 961
Db 1116 CAGGCACTGACGGCCCATCTGTCTCCCTTTCTTGTATGAGGAGCTCATCCAGCCACCC 1175
Qy 962 CACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGGAGAGATGCTTGACATAC 1021
Db 1176 CACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGGAGAGATGCTTGACATAC 1235
Qy 1022 TGCTGCAGGGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGACAGAGTGCCTAAC 1081
Db 1236 TGCTGCAGGGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGACAGAGTGCCTAAC 1295
Qy 1082 AGGGGAGTCCACGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGACACCT 1141
Db 1296 AGGGGAGTCCACGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGACACCT 1355
Qy 1142 GCCAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGCATGTGTGGGCA 1201
Db 1356 GCCAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGCATGTGTGGGCA 1415
Qy 1202 TCGTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCT 1261
Db 1416 TCGTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCT 1475
Qy 1262 CAGCCTATCTCAATCGATCTACAATGCTCTGGAAGGCTGAGCTG 1305
Db 1476 CAGCCTATCTCAATCGATCTACAATGCTCTGGAAGGCTGAGCTG 1519

Search completed: June 7, 2004, 23:47:00

Job time : 579 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 23:21:59 ; Search time 3766 Seconds
(without alignments)
10347.888 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305
Sequence: 1 atggatcctgacagtgatca.....atgtctggaagctgagctg 1305

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_nam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gsl1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1009	77.3	1373	11	BC004855
2	993.4	76.1	995	29	AV420489
3	835.6	64.0	995	29	AV420490
4	805.8	61.7	2177	11	AK078890

5	716.6	54.9	995	29	AY420491	AY420491 Mus muscu
6	699.8	53.6	844	10	BE531100	BE531100 601278466
7	634.2	48.6	975	12	EG288427	EG288427 602388031
8	598.4	45.9	895	13	BU522841	BU522841 AGENCOURT
9	587.8	45.0	617	12	BM795149	BM795149 K-EST0076
10	584.6	44.8	920	10	BE616663	BE616663 601278696
11	551.2	42.2	804	12	EG822314	EG822314 602727917
12	538.2	41.2	569	12	BM686492	BM686492 UI-B-CRO-
13	537.2	41.2	1007	12	EG386903	EG386903 602454702
14	528.2	40.5	956	13	BU523218	BU523218 AGENCOURT
15	513.6	39.4	787	10	BE615750	BE615750 601279885
16	502.8	38.5	879	10	BE616186	BE616186 601278758
17	495	37.9	548	14	CB159713	CB159713 K-EST0219
18	434.2	37.9	503	10	BE514663	BE514663 601317136
19	488.8	37.5	742	12	EG468475	EG468475 602510448
20	481.4	36.9	498	12	BM795128	BM795128 K-EST0076
21	479.8	36.8	744	12	BM008802	BM008802 603618532
22	479.4	36.7	515	14	CB142902	CB142902 K-EST0196
23	474.8	36.4	935	10	BE615106	BE615106 601280475
24	457.6	35.1	920	13	BQ922656	BQ922656 AGENCOURT
25	456.2	35.0	931	13	BU157248	BU157248 AGENCOURT
26	453.8	34.8	977	10	BE614660	BE614660 601281614
27	442.2	33.9	689	12	EG966811	EG966811 602834306
28	440.8	33.8	659	9	AI924527	AI924527 wn51c07.x
29	439.4	33.7	648	9	AI924182	AI924182 wn53f11.x
30	437.2	33.5	645	10	BF118991	BF118991 601755351
31	436.4	33.4	1105	10	BE736425	BE736425 601307316
32	433.2	33.2	482	12	EG984172	EG984172 IL5-CN006
33	428.4	32.8	803	14	CA311909	CA311909 UI-CF-FNO
34	419.8	32.2	640	12	EG822651	EG822651 602725762
35	406	31.1	980	10	BE745289	BE745289 601574064
36	383.4	29.4	936	12	EG387113	EG387113 602455825
37	381.8	29.3	1088	10	BE616293	BE616293 601281476
38	373.4	28.6	408	12	BM762292	BM762292 K-EST0043
39	368.8	28.3	435	10	BF811390	BF811390 CM2-CI017
40	360.2	27.6	850	12	B1251465	B1251465 602994002
41	355	27.2	584	9	AI597567	AI597567 tr92a03.x
42	353.8	27.1	693	12	EG481239	EG481239 602528566
43	353	27.0	711	10	BE937482	BE937482 RC4-ST027
44	352	27.0	644	12	BM791640	BM791640 K-EST0071
45	344	26.4	721	13	BU600097	BU600097 AGENCOURT

ALIGNMENTS

RESULT 1
BC004855
LOCUS Homo sapiens, Similar to transmembrane protease, serine 4, clone
DEFINITION BC004855. 1773 bp mRNA linear HTC 12-JUL-2001
ACCESSION BC004855
VERSION BC004855.1 GI:14709533
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1973)
AUTHORS Strausberg R.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgaps@emil.nih.gov
Tissue Procurement: DCID/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>

contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: p Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8347148
This clone has the following problem: frame shifted.

FEATURES

source

1. 1973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3835263"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH MCC_9"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"

ORIGIN

Query Match 77.3%; Score 1009; DB 11; Length 1973;
Best Local Similarity 89.0%; Pred. No. 1.4e-236;
Matches 1162; Conservative 0; Mismatches 0; Indels 143; Gaps 1;
1 ATGGATCTGACAGTGATCAACTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCC 60
243 ATGGATCTGACAGTGATCAACTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCC 302
61 CGTATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
303 CGTATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 362
121 CTGGGAGTATCATATCTGTGTCTCATCAAGTGAATCTGGATAAATACTACTTC 180
363 CTGGGAGTATCATATCTGTGTCTCATCAAGTGAATCTGGATAAATACTACTTC 422
181 CTCTGGGCGAGCTCTCCACATTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC 240
423 CTCTGGGCGAGCTCTCCACATTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC 482
241 TGTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCCAGTG 300
483 TGTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCCAGTG 542
301 GCAGTCCGCTCTCCAGAGCAGTCCACACTGCAGGTGCTGGACTCGGCCACAGGAA 360
543 GCAGTCCGCTCTCCAGAGCAGTCCACACTGCAGGTGCTGGACTCGGCCACAGGAA 602
361 TGCTTCTCTGCTGTGTTCCAGAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCG 420
603 TGCTTCTCTGCTGTGTTCCAGAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCG 662
421 ATGGGCTACAGAGCAAAACCCACTTCAGAGCTGTGGAGATTGGCCCGACACAGATCTG 480
663 ATGGGCTACAGAGCAAAACCCACTTCAGAGCTGTGGAGATTGGCCCGACACAGATCTG 722
481 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCC 540
723 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCC 782
541 TGTCTCTAGGCTCCCTGTCTCCCTGCACTGTCTGCTGTGGAGAGCTTGAAGACC 600
783 TGTCTCTAGGCTCCCTGTCTCCCTGCACTGTCTGCTGTGGAGAGCTTGAAGACC 842
601 CCCCGTGTGTGGTGGGAGAGGCTCTGTGGATTCTTGGCTTGGAGGTGAGCATC 660
843 CCCCGTGTGTGGTGGGAGAGGCTCTGTGGATTCTTGGCTTGGAGGTGAGCATC 902
661 CAGTAGCAAAACAGCACCTCTGTGGAGGAGCATCTTGAGCCCCCTGAGTCTCTCAG 720
903 CAGTAGCAAAACAGCACCTCTGTGGAGGAGCATCTTGAGCCCCCTGAGTCTCTCAG 962

QY 721 GCAGCCCACTGCTTTCAGAAACATACCATGATGTGTTCAACTGGAGTGGCGGAGGTCA 780
Db 963 GCAGCCCACTGCTTTCAGAAACATACCATGATGTGTTCAACTGGAGTGGCGGAGGTCA 1022
QY 781 GACAACTGGGAGCTTCCCATCCCTGGCTGGCCAGATCATCATCATGATTAATCAAC 840
Db 1023 GACAACTGGGAGCTTCCCATCCCTGGCTGGCCAGATCATCATCATGATTAATCAAC 1082
QY 841 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCACTTC 900
Db 1083 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCACACTCACTTC 1142
QY 901 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 960
Db 1143 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1202
QY 961 CCACCTCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATA 1020
Db 1203 CCACCTCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATA 1243
QY 1021 CTGCTGAGGCGTCACTCCAGGTCATTTACAGCACACAGGTCGAATGCAGACGATGGGTAC 1080
Db 1244 ----- 1243
QY 1081 CAGGGGAAGTCACCCGAGAGATGATGTGTGACGGCATCCCGAAGGGGTGTGGACACC 1140
Db 1244 ----- 1243
QY 1141 TGCAGGCTGACAGTGTGGGCCCCCTGATGATCAATCTGACCACTGGCATGTGGTGGGC 1200
Db 1244 ----AGGTCACAGTGTGGGCCCCCTGATGATCAATCTGACCACTGGCATGTGGTGGGC 1299
QY 1201 ATGCTTACCTGGGCTGTGCTGGGGGCCCCGAGACCCGAGGATATACCAAGGTC 1260
Db 1300 ATGCTTACCTGGGCTGTGCTGGGGGCCCCGAGACCCGAGGATATACCAAGGTC 1359
QY 1261 TCAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1305
Db 1360 TCAGCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1404

RESULT 2

AY420489

LOCUS

DEFINITION

Homo sapiens TMRSS4 gene, VIRTUAL TRANSCRIPT, partial sequence,

genomic survey sequence.

ACCESSION

AY420489.1

VERSION

GI:39776446

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 995)

AUTHORS

Clark, A.G., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,

Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,

Rockville, MD 20850, USA

COMMENT

This sequence as made by sequencing genomic exons and ordering them based on alignment.

AY420489 995 bp DNA linear GSS 17-DEC-2003

Homo sapiens TMRSS4 gene, VIRTUAL TRANSCRIPT, partial sequence,

genomic survey sequence.

AY420489.1 GI:39776446

GSS.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 995)

Clark, A.G., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,

Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Inferring nonneutral evolution from human-chimp-mouse orthologous

gene trios

Science 302 (5652), 1960-1963 (2003)

14671302

2 (bases 1 to 995)

Clark, A.G., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,

Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,

Rockville, MD 20850, USA

This sequence as made by sequencing genomic exons and ordering them based on alignment.

FEATURES
source
Location/Qualifiers
1..995
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..995
/gene="TMPRSS4"
/locus_tag="HCM7241"

ORIGIN
Query Match 76.11%; Score 993.4; DB 29; Length 995;
Best Local Similarity 99.9%; Pred. No. 6.4e-233;
Matches 994; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 152 TCAAGGTGATCTGATATAAATACTACTTCTCTGGGAGGAGCTTCCACTTTCATCCCGA 211
DB 1 TCAAGGTGATCTGATATAAATACTACTTCTCTGGGAGGAGCTTCCACTTTCATCCCGA 60

QY 212 GGAACAGCTGTGTACGAGAGCTGGACTGTCTTCTGGGGAGGAGCAGGAGCACTGTG 271
DB 61 GGAACAGCTGTGTACGAGAGCTGGACTGTCTTCTGGGGAGGAGCAGGAGCACTGTG 120

QY 272 TCAAGAGCTTCCCGAAGGCTGAGTGGAGTCCGCTCTCCAGGACCGATCCACAC 331
DB 121 TCAAGAGCTTCCCGAAGGCTGAGTGGAGTCCGCTCTCCAGGACCGATCCACAC 180

QY 332 TGCAGGTGTGACTCGGCCACAGGAACTGGTCTCTGCTGTCTTTCGACAACTTCACAG 391
DB 181 TGCAGGTGTGACTCGGCCACAGGAACTGGTCTCTGCTGTCTTTCGACAACTTCACAG 240

QY 392 AAGCTCTGCTGAGACAGCTGTAGGAGATGGGTACGAGCAAAACCCACTTTCAGAG 451
DB 241 AAGCTCTGCTGAGACAGCTGTAGGAGATGGGTACGAGCAAAACCCACTTTCAGAG 300

QY 452 CNGTGGAGATCGCCAGACAGGATCTGGATGTTGTTGAATACACAGAAACAGCCACAG 511
DB 301 CNGTGGAGATCGCCAGACAGGATCTGGATGTTGTTGAATACACAGAAACAGCCACAG 360

QY 512 AGCTTCGATCGGAACTCAAGTGGGCTGTCTCTACAGGCTCTGTCTCCCTGCACT 571
DB 361 AGCTTCGATCGGAACTCAAGTGGGCTGTCTCTACAGGCTCTGTCTCCCTGCACT 420

QY 572 GTCCTGCTGTGGAGAGCTGAGACCCCGGTGGTGGTGGGAGGAGGCTCTG 631
DB 421 GTCCTGCTGTGGAGAGCTGAGACCCCGGTGGTGGTGGGAGGAGGCTCTG 480

QY 632 TGGATCTTGGGCTTGGCAGTCCAGATCCAGTACGACAAACAGCACGCTGTGTGAGGGA 691
DB 481 TGGATCTTGGGCTTGGCAGTCCAGATCCAGTACGACAAACAGCACGCTGTGTGAGGGA 540

QY 692 GCATCTGAGACCCCACTGGTCTTCAAGGAGCCCACTGCTTCAGGAACATACCGATG 751
DB 541 GCATCTGAGACCCCACTGGTCTTCAAGGAGCCCACTGCTTCAGGAACATACCGATG 600

QY 752 TGTTCACCTGGAAGTGGGAGGCTCAGACAACTGGGAGCTTCCCATCCCTGGGCTG 811
DB 601 TGTTCACCTGGAAGTGGGAGGCTCAGACAACTGGGAGCTTCCCATCCCTGGGCTG 660

QY 812 TGGCCAAAGATCATCATTAATTAATCAACCCCATGTACCCCAAGACATGATCGGCC 871
DB 661 TGGCCAAAGATCATCATTAATTAATCAACCCCATGTACCCCAAGACATGATCGGCC 720

QY 872 TCATGAAGCTGAGTTCCTCCTCAGGACAGTCCAGGCTCTCTCTGCTGCTG 931
DB 721 TCATGAAGCTGAGTTCCTCCTCAGGACAGTCCAGGCTCTCTCTGCTGCTG 780

QY 932 TCTTTGATGAGGAGCTCACTCCAGCAACCCCACTCTGGATCATTTGATGGGCTTTACGA 991
DB 781 TCTTTGATGAGGAGCTCACTCCAGCAACCCCACTCTGGATCATTTGATGGGCTTTACGA 840

QY 992 ACAGAAATGAGGAGAGTGTCTGACATCTCTGAGGCTCAGTCCAGGCTTCATGCA 1051
DB 841 ACAGAAATGAGGAGAGTGTCTGACATCTCTGAGGCTCAGTCCAGGCTTCATGCA 900

QY 1052 GCACACGGTGAATGACAGAGCTGCTACAGGGGGAAGTCCAGGAGAGATGATGTG 1111
DB 901 GCACACGGTGAATGACAGAGCTGCTACAGGGGGAAGTCCAGGAGAGATGATGTG 960

QY 1112 CAGGCATCCCGAAGGAGGAGTGTGACACTTCCAG 1146
DB 961 CAGGCATCCCGAAGGAGGAGTGTGACACTTCCAG 995

RESULT 3
AY420490 995 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes TMPRSS4 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY420490
VERSION AY420490.1 GI:39776447
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 995)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
Gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 995)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source 1..995
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..995
/gene="TMPRSS4"
/locus_tag="HCM7241"

ORIGIN
Query Match 64.08; Score 835.6; DB 29; Length 995;
Best Local Similarity 84.2%; Pred. No. 3.9e-194;
Matches 838; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 152 TCAAGGTGATCTGATATAAATACTACTTCTCTGGGAGGAGCTTCCACTTTCATCCCGA 211
DB 1 TCAAGGTGATCTGATATAAATACTACTTCTCTGGGAGGAGCTTCCACTTTCATCCCGA 60

QY 212 GGAACAGCTGTGTACGAGAGCTGGACTGTCTTCTGGGGAGGAGCAGGAGCACTGTG 271
DB 61 GGAACAGCTGTGTACGAGAGCTGGACTGTCTTCTGGGGAGGAGCAGGAGCACTGTG 120

QY 272 TCAAGAGCTTCCCGAAGGCTGAGTGGAGTCCGCTCTCCAGGACCGATCCACAC 331
DB 121 TCAAGAGCTTCCCGAAGGCTGAGTGGAGTCCGCTCTCCAGGACCGATCCACAC 180

QY 332 TGCAGGTGTGACTCGGCCACAGGAACTGGTCTCTGCTGTCTTTCGACAACTTCACAG 391
DB 181 TGCAGGTGTGACTCGGCCACAGGAACTGGTCTCTGCTGTCTTTCGACAACTTCACAG 240

QY 392 AAGCTCTGCTGAGACAGCTGTAGGAGATGGGTACGAGCAAAACCCACTTTCAGAG 451
DB 241 AAGCTCTGCTGAGACAGCTGTAGGAGATGGGTACGAGCAAAACCCACTTTCAGAG 300

Db 241 NNNNTCTCGCTGAGACAGCCTGTAGGACAGATGGCTACAGCAGCAACCACTTTTCAGAG 300
QY 452 CTGTGAGATGGCCAGACAGCAGATCTGATGTTGTTGAATCACAGAAAACAGCCAGG 511
Db 301 CTGTGAGATGGCCAGACAGCAGATCTGATGTTGTTGAATCACAGAAAACAGCCAGG 360
QY 512 AGCTTCGCTGCGGAACCTCAAGTGGCCCTGTCTCTCAGGCTCCCTGGTCTCCCTGCAC 571
Db 361 AGCTTCACATCGGAACCTCAAGTGGCCCTGTCTCTCAGGCTCCCTGGTCTCCCTGCAC 420
QY 572 GTCTTCGCTGCGGAAGAGCCTGAAGACCCCTGTGTGTGGTGGGAGGAGGCTCTG 631
Db 421 GTCTTCGCTGCGGAGAGCCTGANNCCCTGTGTGTGGTGGGAGGAGGCTCTG 480
QY 632 TGGATCTTGGCTTGGCAGTCCAGTCCAGTACGACAAACAGCAGCAGTCTGTGGAGGGA 691
Db 481 TGGATCTTGGCTTGGCAGTCCAGTCCAGTACGACAAACAGCAGCAGTCTGTGGAGGGA 540
QY 692 GCATCTCGACCCCACTGGCTCTCACGGCAGCCCACTGCTTCAGGAAACATACCGATG 751
Db 541 GCATCTCGACCCNN 600
QY 752 TGTTCACCTGGAAGTGGCGGAGGCTCAGCAAACTGGCGAGCTTCCCATCCCTGGCTG 811
Db 601 TGTTCACCTGGAAGTGGCGGAGGCTCAGCAAACTGGCGAGCTTCCCATCCCTGGCTG 660
QY 812 TGGCCAGATCATCATCTGATTCAGTCAACCCCACTGATCCCAAGCAATGATCGGCC 871
Db 661 TGGCCAGATCATCATCTGATTCAGTCAACCCCACTGATCCCAAGCAATGATCGGCC 720
QY 872 TGTGAAGTGCAGTTCCTCCACTCACTTCTCAGGCACAGTCAAGCCCATCTCTGCTG 931
Db 721 TCATGAAGTGCAGTTCCTCCACTCACTTCTCAGGCACAGTCAAGCCCATCTCTGCTG 780
QY 932 TCTTTCATGAGGAGTCACTCCAGCCACCCCACTCTGGATCATTTGATGGGCTTTTACGA 991
Db 781 TCTTTCATGAGGAGTCACTCCAGCCACCCCACTCTGGATCATTTGATGGGCTTTTACGA 840
QY 992 AGCAGATCGGAGGAAGTGTCTGATCATCTCTCAGCCGCTGATCCAGTCAATGACA 1051
Db 841 AGCAGATCGGAGGAAGTGTCTGATCATCTCTCAGCCGCTGATCCAGTCAATGACA 900
QY 1052 GCACACGTCGATGACAGACGATCGTACAGGGGAGTCAACGAGATGATGCTGTG 1111
Db 901 GCACACGTCGATGACAGACGATCGTACAGGGGAGTCAACGAGATGATGCTGTG 960
QY 1112 CAGGCATCCCGAAGGGGTGTGGACACCTGCCAG 1146
Db 961 CAGGCATCCCGAAGGGGTGTGGACACCTGCCAG 995

RESULT 4
AK078890
LOCUS
DEFINITION
Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:903062G02 product:SIMILAR TO TRANSMEMBRANE PROTEASE, SERINE 4 homolog [Mus musculus], full insert sequence.
AK078890
VERSION
AK078890.1 GI:26098158
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Oneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2177)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers
1. 2177
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:903062G02"
/db_xref="MGI:2397607"
/db_xref="taxon:10090"
/clone="903062G02"
/sex="male"
/tissue_type="colon"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 304..1507
/notes="SIMILAR TO TRANSMEMBRANE PROTEASE, SERINE 4 homolog

```
[Mus musculus] (SPTR|AAH21368, evidence: FASTY, 92%ID,
100%length, match=1201)
putative"
2154..2159
/note="putative"
2177
/note="putative"

polyA_signal
polyA_site

ORIGIN
Query Match      61.7%; Score 805.8; DB 11; Length 2177;
Best Local Similarity 82.7%; Pred No. 1.3e-185;
Matches 921; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 193 CCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAGTGTCTCTGGG 252
Db CCCCTGACCTTCATTACAGAGGGGCCAGTTGTGTGACGGCCACCTTACTGCGCCTCAGG 451

QY 253 GAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGGCAGTCCGCTC 312
Db GAGGATGAGGAACACTGTGTCAAGAGCTTCCCTGAAAAGCCCGAGTGGCAGTCCGGCTC 511

QY 313 TCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAACCTGGTTCTCTGCC 372
Db TCCAGGACAGATCCACCTGCAGGTGTGGATGCAGCCACAGGGAACCTGGGCTCAGTC 571

QY 373 TGTTTCGACACTTCACAGAGCTCTCGTGAGACAGCTGTAGGCAGATGGCTACAGC 432
Db TGTTTCGACACTTCACAGAGCACTGGCCAGACAGCTGCAGACAGATGGCTATGAC 631

QY 433 AGCAAAACCCATTTCAGAGCTGTGAGATGTCGCCACAGACAGGATCTGGATGTGTGAA 492
Db AGCCAGCCCGTTTCAGAGCAGTGGAGATCCGTCCAGATCAGAACCTCCCTGTGCTCAA 691

QY 493 ATCAGAGAAACAGCCAGGAGCTTCGCATGCGGAATCAAGTGGGCGCTGTCTTCAGGC 552
Db GTCACAGGAAACAGCCAGGAACCTTCAGGTGCAGAAATGGAAGCAGATCTGCTCTCAGGC 751

QY 553 TCCCTGGTCTCCCTGCACCTGTCTCCCTGTGGAGAGCGCTGAGACCCCGCGTGGTG 612
Db TCCCTGGTTCCTTGGCGTCCCTTGAAGAGAGCGCTGGAAGACCTCTCGTGGTG 811

QY 613 GGTGGGAGGAGGCGCTCTGTGATCTTTGGCTTGGCAGGTTCAGATCAGATCAGTACGACAAA 672
Db GGTGGGTGGAGGCGCTCTGTGATCTTTGGCGTGGCAGGTTCAGATCAGTACGACAA 871

QY 673 CAGCAGCTGTGTGAGGAGGAGCTCTGACCCCGCTGCTGCTCAGCGAGCCACTGC 732
Db CAGCATGTCTGTGTGGGAGCATCTGTGATCCCGCTGCTGCTCAGCAGCCCACTGC 931

QY 733 TTCAGGAAACATACCGATGTGTTCACCTGGAAGGTGCGGCGAGGCTCAGACAACTGGGC 792
Db TTCAGGAGATATCTTGATGTGTCAAGCTGGAAGGTTCAGGCGAGGCTCAACATCTGGT 991

QY 793 AGCTTCCCATCCCTGGCTGTGGCCAGAGATCATCATCATTTGAATTCACCCCATGTACCCC 852
Db AACTCTCCATCCCTGTGCTGGCCAGAGATCTTCATCGTGAACCCCAATCCTCTGTACCCC 1051

QY 853 AAAGCAATGATCAGCCCTCATGAGCTGAGCTGCGAGTTCGCCACTCAGTCTTCAGGCACTGC 912
Db AAAGAGAGCAATTCCTCTGTTAAGCTGAGATGCGCACTCAGTCTTCAGGCTCAGTC 1111

QY 913 AGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCACTCTGGATC 972
Db AGGCCCATCTGTCTGCCCTTCTTGTATGAGGTGCTTGTCCAGCCACACCACTCTGGGTC 1171

QY 973 ATGATGGGGCTTTACGAGAGCAGATGGAGGAGAGATGCTGATATCTGTCGAGGG 1032
Db ATGATGGGGCTTTACAGAGAAACCGAGAGAAAGATGCTGATGATCTACTGCGAGCA 1231

QY 1033 TCAGTCCAGGTCAATTCAGACAGCAGCGGTGCAATGCAGACGATGCGGTACGAGGGGAAGTC 1092
Db TCAGTCCAGGTCAATTCAGACAGCAGCGGTGCAATGCAGAGGATGCGGTACGAGGGGAAGTC 1291
```

```
QY 1093 ACCGAGAAATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTGCGAGGTGAC 1152
Db ACCGCTGAGATGCTGTGTGAGGTACCCACAGGGTGGCAAGGACACCTGCCAGGTGAC 1351

QY 1153 ASTGTTGGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1212
Db AGTGGTGGGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1411

QY 1213 GGTATGCTGCGGGGCGCGGAGCACCAGGAGTATACACCAAGGTCTCAGCTATCTC 1272
Db GGCCATGATGCGGGCGCGGAGTACTCTGTGAGTGTATACCAAGGTCACTGCTATCTC 1471

QY 1273 AACTGATCTACAAATGCTGTGGAAGGTGAGCTG 1305
Db AACTGATCTACAAATGCTGTGGAAGGTGAGATG 1504

RESULT 5
AY420491 995 bp DNA linear GSS 17-DEC-2003
LOCUS Mus musculus TWPRSS4 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY420491
VERSION AY420491.1 GI:39776448
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 995)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 995)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
TITLE Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
JOURNAL
COMMENT This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source
1..995
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..5995
/gene="TWPRSS4"
/locus_tag="HCM7241"

gene

ORIGIN
Query Match 54.9%; Score 716.6; DB 29; Length 995;
Best Local Similarity 82.5%; Pred No. 6.9e-165;
Matches 821; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 152 TCAAGGTGATTCGATAAATACTACTTCTCTGCGGCGAGCTCTCCACTTCATCCCGA 211
Db TCAAGGTGATTCGATAAATACTACTTCTCTGCGGCGAGCTCTCCACTTCATTCAGA 60

QY 212 GGAAGCAGCTGTGTGACGGAGAGCTGAGCTGTCCTTGGGGGAGGAGGAGGAGCAGCTGTG 271
Db GGGGCCAGTGTGTGACGGCCACCTTGACTGCGCCTCAGGGGAGGATGAGGAACACTGTG 120

QY 272 TCAAGAGCTTCCCGAAGGCGCTGAGTGGCAGTCCCGCTCTCCAGGACCGGATCCACAC 331
Db TCAAGAGCTTCCCGTAAAGCCCGAGTGGCAGTCCGCTCTCCAGGACGAGATCCACCC 180
```

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM266 row: d column: 18
High quality sequence stop: 712.
Location/Qualifiers
1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610337"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_39"
/notes="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

332 TGCAAGTGTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAG 391
Db
181 TGCAAGTGTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAG 240
Qy
392 AGCTCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAG 451
Db
241 AGCACTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAG 300
Qy
452 CTGTGAGATTTGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 511
Db
301 CAGTGGAGATTCGTCAGATCAGAACTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 360
Qy
512 AGCTCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 571
Db
361 AACTTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 420
Qy
572 GTCTTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 631
Db
421 GCCTTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 480
Qy
632 TGAATCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 691
Db
481 TGAATCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 540
Qy
692 GCATCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 751
Db
541 GCATCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 600
Qy
752 TGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 811
Db
601 TGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 660
Qy
812 TGGCCAGATCATCATCTGATTTCAACCCCATGATGACCCCAAGACATGACATGACATGACATG 871
Db
661 TGGCCAGATCATCATCTGATTTCAACCCCATGATGACCCCAAGACATGACATGACATGACATG 720
Qy
872 TCATGAGTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 931
Db
721 TGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 780
Qy
932 TCTTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 991
Db
781 TCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 840
Qy
992 AGCAGATCTGAGGAACTGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1051
Db
841 AGAAGACGAGGAACTGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 900
Qy
1052 GCACAGTCTGATGAGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1111
Db
901 GCACAGTCTGATGAGTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 960
Qy
1112 CAGGCACTCCGGAAGGGGTGGACACCTGCCAG 1146
Db
961 CAGGCACTCCGGAAGGGGTGGACACCTGCCAG 995

RESULT 6
BE531100
LOCUS
DEFINITION
601278466F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610337 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 844)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM266 row: d column: 18
High quality sequence stop: 712.
Location/Qualifiers
1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610337"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_39"
/notes="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

Query Match 53.6%; Score 699.8; DB 10; Length 844;
Best Local Similarity 96.4%; Pred. No. 8.4e-161;
Matches 749; Conservative 0; Mismatches 22; Indels 6; Gaps 3;

Qy 118 AGCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 177
Db 1 AGCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 60
Qy 178 TTCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 237
Db 61 TTCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 120
Qy 238 GACTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 297
Db 121 GACTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 180
Qy 298 GTGCGAGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 357
Db 181 GTGCGAGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 240
Qy 358 AACTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 417
Db 241 AACTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 300
Qy 418 CAGATGGGCTACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCACAGGAACTTTCAG 477
Db 301 CAGATGGGCTACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCACAGGAACTTTCAG 360
Qy 478 CTGATGTTCTGTTGAATTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAG 537
Db 361 CTGATGTTCTGTTGAATTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAGAACTTCAG 420
Qy 538 CCTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 597
Db 421 CCTGTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 480
Qy 598 ACCCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 657
Db 481 ACCCTCTCTGAGTCTGGCCACAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAG 540
Qy 658 ATCCAGTACGAACTTCAGGAACTGTCTCTCTGCTTTGCAAACTTTCACAGGAACTTTCAGGAACTTTCAG 717

Db 541 ATCCAGTAGCAGCAACAGCAGCTCTGTGGAGGGAGCATCTTGGAGCCCCCACTGGGTCCTC 600
 Qy 718 ACGGAGGCCACTGTCTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGCAGGC 777
 Db 601 ACGGAG-CCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGCAGGC 659
 Qy 778 TCAGACAACACTGGGAGCTCCCTCCCTGGCTGTGGCCAGATCATCATATTGAATTC 837
 Db 660 TCAGAC-AACTGGGAGCTTCCATCCCTGGGCTGTGGCCAGATCATCATATTGAATTC 718
 Qy 838 AACCCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTC 894
 Db 719 AACC-----TGTTCCCAAGACATGACATCGCCCTCATGAAGCTGCGTTCCTCCCTCAC 771

 RESULT 7
 BG288427
 LOCUS
 DEFINITION 602388091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5', linear EST 21-FEB-2001
 mRNA sequence.
 ACCESSION BG288427.1 GI:13043459
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 975)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10409 row: e column: 23
 High quality sequence stop: 659.
 Location/Qualifiers
 1..975
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone=IMAGE:4516990"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 93"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

 ORIGIN
 Query Match 48.6%; Score 634.2; DB 12; Length 975;
 Best Local Similarity 96.4%; Pred. No. 1.2e-144;
 Matches 560; Conservative 0; Mismatches 23; Indels 2; Gaps 1;

 Qy 584 GGAAGAGCTGGAAGACCCCTGTGTGGTGGGGAGAGCCCTCTGTGATCTTGGC 643
 Db 1 GGAAGAGCTGGAAGACCCCTGTGTGGTGGGGAGAGCCCTCTGTGATCTTGGC 60

 Qy 644 CTTCAGCTCAGCATCCAGTACGACAAACACACGCTGTGTGGAGGAGCATCTGGACC 703
 Db 61 CTTCAGCTCAGCATCCAGTACGACAAACACACGCTGTGTGGAGGAGCATCTGGACC 120

 Qy 704 CCACCTGGGTCTCAGCGAGCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGA 763
 Db 121 CCACCTGGGTCTCAGCGAGCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGA 180

Qy 764 AGTGGGGGAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCAAAGATCA 823
 Db 181 AGTGGGGGAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCAAAGATCA 240

 Qy 824 TCATCATTTGAATCAACCCCATGTACCCAAAGACATGACATCGCCCTCATGAAGTGC 883
 Db 241 TCATCATTTGAATCAACCCCATGTACCCAAAGACATGACATCGCCCTCATGAAGTGC 300

 Qy 884 AGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCGCTTTCTTTGATGAGG 943
 Db 301 AGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCGCTTTCTTTGATGAGG 360

 Qy 944 AGTCACTCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTAGGAAGCAATGGAG 1003
 Db 361 AGTCACTCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTAGGAAGCAATGGAG 420

 Qy 1004 GGAAGATGTCTGACATACTCTGTCAGCGCTCAGTCCAGGTCAATTGACAGCACACGGTGCA 1063
 Db 421 GGAAGATGTCTGACATACTCTGTCAGCGCTCAGTCCAGGTCAATTGACAGCACACGGTGCA 480

 Qy 1064 ATGCAGACGATGGTACCAGGGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCCG 1123
 Db 481 ATGCAGACGATGGTACCAGGGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCCG 540

 Qy 1124 AAGGGGGTGTGGACACCTGCCAGGTGACAGTGTGGGGCCCTGATGTACCAATCTGACC 1183
 Db 541 AAGGGGGTGTGGACACCTGCCAGGTGACAGTGTGGGGCCCTGATGTACCAATCTGACC 600

 Qy 1184 AGTGGCATGTGTGGGATCGTTAGTGGGCTATGGCTGCGGGGCGCCGAGCACCCCGAG 1243
 Db 601 AGTGGCATGTGTGGGATCGTTAGTGGGCTATGGCTGCGGGG--CCGAGCACCCCGAG 658

 Qy 1244 GAGTATACACCAAGTCTCAGCCTA 1268
 Db 659 AGTTACCACCAAGTCTCAAGCTA 683

 RESULT 8
 BUS22841
 LOCUS
 DEFINITION AGENCOURT_10158166 NCI_CGAP_Co24 Mus musculus cDNA clone
 IMAGE:6529422 5', mRNA sequence.
 ACCESSION BUS22841
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14129 row: i column: 06
 High quality sequence stop: 665.
 Location/Qualifiers
 1..895
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone=IMAGE:6529422"
 /lab_host="DH10B (T1 phage-resistant)"

FEATURES
source

```
ORIGIN
Query Match 45.9%; Score 598.4; DB 13; Length 895;
Best Local Similarity 83.0%; Pred. No. 7.3e-136;
Matches 694; Conservative 0; Mismatches 141; Indels 1; Gaps 1;

/clone lib="NCI CGAP Co24"
/Note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

QY 454 GTGCGATTGGCCCGACAGCAGATCGGATGTGTGTGAATCACAGAAACAGCCAGAG 513
Db 17 GGTGGGTCCGTCAGATCAGAACCTCCCTGTGTCTCAAGTCACAGGAACAGCCAGAA 76
QY 514 CTTGCGATCGGAACCTCAAGTGGGCCCTCTCTCAGGCTCCCTGTCTCCCTGCACTGT 573
Db 77 CTTGAGTGCAGAAATGGAAGCAGATCTCTGCTCTCAGGCTCCCTGTCTCCCTGCTGC 136
QY 574 CTTGCTGTGGGAAGCCTGAGACCCCTGTGTGTGTGGTGGGAGGAGCCTCTGTG 633
Db 137 CTTGACTGTGGAAGAGCCTGAGACTCTCTGTGTGTGGTGGGAGGAGCCTCTGTG 196
QY 634 GATTCCTTGCCCTTGGCAGCTCAGCATCCAGTACGACAAACAGCAGCTCTGTGGAGGAGC 693
Db 197 GATTCCTTGCCCTTGGCAGCTCAGCATCCAGTACGACAAACAGCAGCTCTGTGGAGGAGC 256
QY 694 ATCTGGAACCCCACTGTGGTCTCTCAGGAGCCCACTGCTTCAAGAAACATACCGATGTG 753
Db 257 ATCTGGATCCCACTGTGGTCTCTCAGGAGCCCACTGCTTCAAGAAACATCTTGTATGTG 316
QY 754 TTCACTGGAAGTGGGAGCAGCTCAGCAAACTGGGAGCTTCCATCCCTGCTGTG 813
Db 317 TCAAGCTGGAAGTGGGAGCAGCTCAGCAAACTGGGAGCTTCCATCCCTGCTGTG 376
QY 814 GCCAAGATCATCATTTGAATTAACCCCATGTATCCCAAGAAACATGATGATGCTC 873
Db 377 GCCAAGATCTCATGCTGAACCCCATCTGTATCCCAAGAAACATGATGATGCTC 436
QY 874 ATGAAGCTCAGTCCCACTGCTTCTCAGCAGCAGTCCAGCCCATCTGTGCTCCTTC 933
Db 437 GTTAAGCTCAGATGCCACTCAATCTCAGGCTCAGTCAGGCCCATCTGTGCTCCTTC 496
QY 934 TTTGATGAGAGTCACTCCAGCCACCCCACTCTGGATCATTTGATGGGCTTTACGAAG 993
Db 497 TCTGATGAGTCTTGTCCAGCCACACCACTGGCTGGCTTGTGATGGGCTTTACGAA 556
QY 994 CAGAATGGAAGAGATGTTGACATATCTGTGAGGGGTGATGTCAGGTATTTGACAGC 1053
Db 557 GAAACCGGAGGAAGATGTTGACATGTTACTTGCAGGATCATCTCAGGTATCATGACAGC 616
QY 1054 ACACGCTGCAATGCAACATCGCTACAGGGGGAAGTCAACGAGAGATGATGTGCA 1113
Db 617 ACACGCTGCAATGCAACATCGCTACAGGGGGAAGTCAACGAGATGATGTGTCGA 676
QY 1114 GGCATCCCGGAAGGGGTGTGACACCTGCGAGGTGACAGTGTGGGCCCTGTATGATAC 1173
Db 677 GGTATCCCAACAGGGTGGCAAGGACACCTGCCAGGGTGAAGTGTGGGCCCTTTGATGATAC 736
QY 1174 CAATCTGACAGTGGCATGTGTGGGATCTGTAGCTGGGCTATGCTGGGGGGCCG 1233
Db 737 CATTTGACAAATGGCAGGTAGTAGGCATCTGTAGCTGGGCGCATGATGGCGGCCCA 796
QY 1234 AGCAC-CCAGGAGTATACCAAGCTCTCAGCTTATCTCAACTGGATCTCAATG 1288
Db 797 AGTACTCCCTGGAGTGTATACCAAGCTCACTGCTTATCTCAACTGGATCTCAATG 852
```

RESULT 9

```
BM795149 617 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0076707 S21SNUS20 Homo sapiens cDNA clone S21SNUS20-27-H07 5',
DEFINITION mRNA sequence.
```

```
BM795149
BM795149.1 GI:19143381
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 617)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 27 row: H column: 07
High quality sequence stop: 617.
Location/Qualifiers
1. 617
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S21SNUS20-27-H07"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/lab_host="Top10F"
/clone_lib="S21SNUS20"
/Note="Organ: Stomach; Vector: pTZ18RPI; Site: 1; EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
```

ORIGIN

```
Query Match 45.0%; Score 587.8; DB 12; Length 617;
Best Local Similarity 98.1%; Pred. No. 2.3e-133;
Matches 605; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 225 TGACGAGAGCTGGATGTCCTTGGGAGGACGAGGAGCAGCTGTGTCAAGAGCTTCCC 284
Db 1 TGACGAGAGCTGGATGTCCTTGGGAGGACGAGGAGCAGCTGTGTCAAGAGCTTCCC 60
QY 285 CGAAGGCTGTAGGCTGCGAGTCCGCTCTCCAGGACCGATCCACTCCAGTGTCTGGA 344
Db 61 CGAAGGCTGTAGGCTGCGAGTCCGCTCTCCAGGACCGATCCACTCCAGTGTCTGGA 120
QY 345 CTCGGCCACAGGGAAGTGTCTCTCCCTGTTTCGACAACTTCACAGAGCTCTCGCTGA 404
Db 121 CTCGGCCACAGGGAAGTGTCTCTCCCTGTTTCGACAACTTCACAGAGCTCTCGCTGA 180
QY 405 GACAGCTGTAGGCTGCGAGTCCGCTCTCCAGGACCGATCCACTTTTCAGAGCTGTGAGATGG 464
Db 181 GACAGCTGTAGGCTGCGAGTCCGCTCTCCAGGACCGATCCACTTTTCAGAGCTGTGAGATGG 240
QY 465 CCCAGCAGGAGTCTGGATGTTGTTGAAATACAGAAAAACAGCCAGGAGCTTCGATGCG 524
```



```

VERSION      BG822914.1  GI:14170501
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    NIH-MGC http://mgi.nci.nih.gov/
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
COMMENT      Email: cgaabs-remail.nih.gov
              Tissue procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: NIH Intramural Sequencing Center
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCW1735 row: b column: 05
              High quality sequence stop: 793.
              Location/Qualifiers
FEATURES     source
             1..804
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4867108"
                /tissue_type="adenocarcinoma cell line"
                /lab_host="PH10B (phage-resistant)"
                /clone_lib="NIH_MGC_15"
                /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGACACAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)"
ORIGIN
Query Match      42.2%; Score 551.2; DB 12; Length 804;
Best Local Similarity 99.0%; Pred. No. 2,7e-124;
Matches 586; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1 ATGGATCTTGACAGTATCAACTCTGAAACAGCTCTGATGTCAAAACCCCTGGCGAAACCC 60
DB 215 ATGGATCTTGACAGTATCAACTCTGAAACAGCTCTGATGTCAAAACCCCTGGCGAAACCC 274
QY 61 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 275 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 334
QY 121 CTGGCGAGTATCATCATTTGTGGTGTCTTCATCAAGGTGANTCTCGATAAATACTACTTC 180
DB 335 CTGGCGAGTATCATCATTTGTGGTGTCTTCATCAAGGTGANTCTCGATAAATACTACTTC 394
QY 181 CTCTGGGGAGAGCTTCACATTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAC 240
DB 395 CTCTGGGGAGAGCTTCACATTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAC 454
QY 241 TGTCCCTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTG 300
DB 455 TGTCCCTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTG 514
QY 301 CGAGTCCGCTCTCCAGAGCCGATCCACATCTGCAGGTCTCGACTCGGCCACAGGAAE 360
DB 515 GCAGTCCGCTCTCCAGAGCCGATCCACATCTGCAGGTCTCGACTCGGCCACAGGAAE 574
QY 361 TGGTCTCTCGCTGTGTTTCAGCAACTTTCACAG-AAGCTCTCGCTGAGACAGCTGTAGGCA 419
DB 575 TGGTCTCTCGCTGTGTTTCAGCAACTTTCACAGAAAGCTCTCGCTGAGACAGCTGTAGGCA 634
QY 420 GATGGGCTACAGCAGCAAAACCCATTTCAGAGCTGTGGAGATTGGCCAGACCAAGTCT 479

```

```

Db 635 GATGGGCTACAGCAGCAAAACCCATTTCAGAGCTGTGGAGATTGGCCAGACCAAGTCT 694
QY 480 GGATGTTGTGTAATCACAGAAACAGCCAGAGCTTCGCATCGCGAACTCAAGTGGGC 539
DB 695 GGATGTTGTGTAATCACAGAAACAGCCAGAGCTTCGCATCGCGAACTCAAGTGGG-C 753
QY 540 CTCTCTCTCAGGCTCCCTGGTCTCCCTGCACACTGTCTTGCTGTGGGAAGC 591
DB 754 CTCTCTCTCAGGCT-CCTGGTCTCCCTGAACCTGTCTTGCTGGGGAGAGC 804

```

```

RESULT 12
BM686292
LOCUS
DEFINITION
UI-E-CRO-adj-g-09-0-UI.r1 UI-E-CRO Homo sapiens cDNA clone
UI-E-CRO-adj-g-09-0-UI 5', mRNA sequence.
BM686292
ACCESSION
BM686292.1 GI:18999550
VERSION
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 569)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1..569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CRO-adj-g-09-0-UI"
/tissue_type="eye anterior segment"
/dev_stage="adult"
/lab_host="UI-E-CRO"
/clone_lib="UI-E-CRO"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CRO is a cDNA library containing the following
tissue(s): eye anterior segment. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA,
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AATGCCGCAAT. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```

```

FEATURES
source

```

```

ORIGIN

```



```
Query Match      41.2%; Score 538.2; DB 12; Length 569;
Best Local Similarity 99.1%; Pred. No. 3.5e-121;
Matches 562; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 517 CGCATGCGAACTCAAGTGGGCCCTGTCTCTCAGGCTGCC-TGGTCTCCCTGCATGCTGT 575
Db 4 CGAGGCGCGAACTCAAGTGGGCCCTGTCTCTCAGGCTGCCCTTGGTCTCCCTGCATGCTGT 63
QY 576 TGCCTGTGGGAAGAGCTCAAGACACCCCGCTGTGTGGTGGGAGGAGGCTCTGTGGA 635
Db 64 TGCCTGTGGGAAGA-CTGGAAGACCCCGCTGTGTGGTGGGAGGAGGCTCTGTGGA 122
QY 636 TTCTTGGCCTTGGCAGGTGAGCATCCAGTACGACAAACAGACACGCTGTGTGGAGGAGCAT 695
Db 123 TTCTTGGCCTTGGCAGGTGAGCATCCAGTACGACAAACAGACACGCTGTGTGGAGGAGCAT 182
QY 696 CCTGACACCCCACTGGGTCTCAAGCGACGCCCACTGCTTCAGGAACATACCGATGTTT 755
Db 183 CTTGACACCCCACTGGGTCTCAAGCGACGCCCACTGCTTCAGGAACATACCGATGTTT 242
QY 756 CAACTGGAAGTGGCGGAGGCTCAGACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGC 815
Db 243 CAACTGGAAGTGGCGGAGGCTCAGACAACTGGGCGAGCTTCCCATCCCTGGCTGTGGC 302
QY 816 CAAATCATCATCATGTAATCAACCCCATGTACCCCAAGACAAATGATGACATCGCCCTCAT 875
Db 303 CAAATCATCATCATGTAATCAACCCCATGTACCCCAAGACAAATGATGACATCGCCCTCAT 362
QY 876 GAAGTGCAGTTCCTCCACTCAGTTCCTCAGGACACGTCAAGGCCCATCTGTCTGCCCTTCT 935
Db 363 GAAGTGCAGTTCCTCCACTCAGTTCCTCAGGACACGTCAAGGCCCATCTGTCTGCCCTTCT 422
QY 936 TGATGAGGAGTCACTCCAGCACCCCACTCTGGATCATTTGGATGGGCTTTTACGAACA 995
Db 423 TGATGAGGAGTCACTCCAGCACCCCACTCTGGATCATTTGGATGGGCTTTTACGAACA 482
QY 996 GAATGAGGAGAGTGTCTGACATCTCTGACGAGCTCAGCCGCTCAGTCCAGTCAATGACAGAC 1055
Db 483 GAATGAGGAGAGTGTCTGACATCTCTGACATCTCTGACGAGCTCAGCCGCTCAGTCCAGTCAATGACAGAC 542
QY 1056 ACGGTGCAATGACAGCATGCGTACCA 1082
Db 543 ACGGTGCAATGACAGCATGCGTACCA 569

RESULT 13
BG386903
LOCUS
DEFINITION
602454702F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582969 5',
mRNA sequence.
ACCESSION
BG386903
VERSION
BG386903.1 GI:13280452
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1007)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M1306 row: c column: 02
High quality sequence stop: 763.
Location/Qualifiers
```

```
source
1..1007
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4582969"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_15"
/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

ORIGIN
Query Match      41.2%; Score 537.2; DB 12; Length 1007;
Best Local Similarity 87.1%; Pred. No. 8.7e-121;
Matches 660; Conservative 0; Mismatches 88; Indels 10; Gaps 6;

QY 1 ATGATCCTCAGAGTATCAACCTCTGACAGAGCTCGATGTCAAACCCCTGGCGAAACCC 60
Db 215 ATGATCCTCAGAGTATCAACCTCTGACAGAGCTCGATGTCAAACCCCTGGCGAAACCC 274
QY 61 CGTATCCCCATGAGACCTTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
Db 275 CGTATCCCCATGAGACCTTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 334
QY 121 CTGCGAGTATCATCATTTGTGGTGTCTCATCAAGGTGATTTGGATAAATACTACTTC 180
Db 335 CTGCGAGTATCATCATTTGTGGTGTCTCATCAAGGTGATTTGGATAAATACTACTTC 394
QY 181 CTCTGGGGGAGGCTCTCCACTTCATCCCGAGAACGAGCTGTGTCACGAGAGCTGGAC 240
Db 395 CTCTGGGGGAGGCTCTCTCCACTTCATCCCGAGAACGAGCTGTGTCACGAGAGCTGGAC 454
QY 241 TGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTCGAGTG 300
Db 455 TGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTCGAGTG 514
QY 301 GCAGTCCGCTCTCCAGGACCCATCCACTCCAGTGTCTGAGCTCGGCTCGGACAGGGAAC 360
Db 515 GCAGTCCGCTCTCCAGGACCCATCCACTCCAGTGTCTGAGCTCGGCTCGGACAGGGAAC 574
QY 361 TGGTTCCTCTCCCTGTTCGACAACTTCACAGAAAGCTCTCGCTGAGACAGCTCTAGGCAG 420
Db 575 TGGTTCCTCTCCCTGTTCGACAACTTCACAGAAAGCTCTCGCTGAGACAGCTCTAGGCAG 634
QY 421 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCGACACAGGATCTG 480
Db 635 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCGACACAGGATCTG 694
QY 481 GATGTTGTTGAAATCA-CAGAAACACGACGAGGAGCTT-CGATGCGGAACTCAAGTGGGC 538
Db 695 GATGTTGTTGACATCACCAAGAAACAGACGAGGAGCTTCGATGCGGAACTCAAGTGGGG 754
QY 539 CCT-GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTG-CCTGTGGGAAGAGCTGAA 596
Db 755 CCTGGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTG-CCTGTGGGAAGAGCTGAA 814
QY 597 GACCCCGCTGTGTGGTGGGAGGAGGCTC-----TGTGATTTCTTGGCTTGGCAG 651
Db 815 GAACCCCGCTGTGTGGTGGGAGGAGGCTC-----TGTGATTTCTTGGCTTGGCAG 674
QY 652 GTCAGCATCCAGTACGACAAACAGACAGCTGTGTG-GAGGAGGAGCATCTCTGACACCCCTG 710
Db 875 AGACACAATCAGTACCAACACACCCAGTCTGTGTGAGAGAGAGACACTCGGGGACCCACG 934
QY 711 GGTCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 748
Db 935 GGCACCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 972
```

RESULT 14	BUS23218	956 bp	mrna	linear	EST 13-SEP-2002
LOCUS	AGENCOURT 10154487 NCI CGAP C024 Mus musculus CDNA clone				
DEFINITION	IMAGE:6523864 5', mRNA sequence.				
ACCESSION	BUS23218				
VERSION	BUS23218.1	GI:22833656			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 956)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: The Cepko Laboratory CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLNL4130 row: k column: 16 High quality sequence stop: 706. Location/Qualifiers				
FEATURES	1..956				
source	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="FVB/N"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:6523864"				
	/lab_host="DH10B (T1 phage-resistant)"				
	/clone_lib="NCI CGAP C024"				
	/note="Organ: Vector; PCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."				
ORIGIN					
Query Match	40.5%; Score 528.2; DB 13; Length 956;				
Best Local Similarity	81.2%; Pred. No. 1.4e-118;				
Matches	662; Conservative 0; Mismatches 148; Indels 5; Gaps 4;				
QY	278 GCTTCCCGAAGGCGCTGAGTGGGAGTCCGCTCTCCAGGACCGATCCACATGACAGG 337				
DB	38 GCGTCGTGAAAAGCCGAGTGGGAGTCCGCTCTCCAGGACGATCCACCTGCGAGG 97				
QY	338 TGTGTGACTCGGCCACAGGGAATGTTCTCTGCTCTGTTTGCACAACTTCACAGAAGTCC 397				
DB	98 TGTGTGATCAGCCACAGGAGCTGGGCTCAGTCTGTTTGCACAACTTCACAGAAGCAC 157				
QY	398 TGTGTGAGACAGCTGTAGGAGATGGGTACAGGACCAACCCACTTCAGAGCTGTGG 457				
DB	158 TGCCCAAGACAGCCTGCACAGATGGGGTATGACAGGACGCGCTTCAGAGAGTGG 217				
QY	458 AGATTGCGCCACAGCAGGATCTGGATGTTGTTGAATCAGAAACAGCCAGGAGCTTC 517				
DB	218 AGATCGCTCAGATCAGACCTCCCTGTTGCTCAAGTACAGGAAACAGCCAGGAATTC 277				
QY	518 GCATGCGGAACCTCAAGTGGGCGCTGTCTCAGGCTCCCTGTCTCCCTGCACCTCTTG 577				
DB	278 AGGTGACAGATGAGACAGATCCTGCCTCTCAGGCTCCCTGTGTTCTTCCTGCGCTCG 337				
QY	578 CTGTGCGGAAGCCCTGAAGACCCCGCTGTGGTGGGAGGAGGCGCTCTGTGGATT 637				
DB	338 ACTGTGGAAGAGCCCTGAAGACTTCCTCTGTTGTTGGTGGGTGGGAGCCCTGTGGATT 397				
QY	638 CTTGGCCTTGGCAGGTTCAGATCCAGTACGACAAACAGACACGCTGTGTGGAGGAGCATCC 697				
Db	398 CTTGGCCGTGGCAGGTTCAGATCCAGTACGACAAACAGACGATCTGTGTGGAGCATCC 457				
QY	698 TGGACCCCACTGGGTCTTCACGGCAGCCCACTGTTTTCAGGAAACATACCGATGTGTCA 757				
DB	458 TGGATCCCACTGGATCTTCACAGAGCCCACTGTTTTCAGGAAAGTATTTGATGTGTCAA 517				
QY	758 ACTGGAAGGTGGGGCAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGCTGTGGGCA 817				
DB	518 GCTGGAAGGTTCAGGCGAGGCTCAAACTACTGGTAACTCTCCATCTTCCTGTGGGCA 577				
QY	818 AGATCATCATCATGTAATTCACCCCATGTATCCCAAGACAAATGACATCGCCCTCATGA 877				
DB	578 AGATCTTCATCTGCTGAACCCATCTCTGTATCCCAAGAGAGACATTCGCCCTTGTGA 637				
QY	878 AGTGTGAGTTCCTCAGTCACTTCTCTCAGGCACAGTTCAGGCCCATCTGTGTGCCCTTCTTG 937				
DB	638 AGTGTGAGATGCCACTCACAATCTCAGGCTCAGTCAGGCCCATCTGTGTGCCCTTCTTG 697				
QY	938 ATGAGAGTCACTCCAGCACCCCACTCTGATCATTTGGATGGGG--CTTTACGAAGCA 995				
DB	698 ATGAGGTGTTTGTCCAGCCACACACCATCTGGGTTCATTTGGATGGGGCTTTACGAAGAA 757				
QY	996 GAATGAGGAGATCTCTGACATATCTGTC-AGGCGTCAGT-CCAGGTCAATTCACACAGC 1053				
DB	758 AACCGAGGAGATGTCCTGACATGCTGCTGATGCTGCAAGGCATCAGTCCAGGTTCATCGAACGC 817				
QY	1054 ACACGGT-GCAATGACAGCATGCGTACCAGGGG 1087				
DB	818 ACACGGTGGCAATGACAGGATGCTTACCAAGGG 852				
RESULT 15	BB615750	787 bp	mrna	linear	EST 24-AUG-2000
LOCUS	601279885F1 NIH_MGC_39 Homo sapiens cdna clone IMAGE:3621980 5',				
DEFINITION	mRNA sequence.				
ACCESSION	BB615750.1	GI:9897349			
VERSION	BB615750.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 787)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LICM296 row: i column: 21 High quality sequence stop: 690. Location/Qualifiers				
FEATURES	1..787				
source	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3621980"				
	/tissue_type="adenocarcinoma"				
	/lab_host="NIH_MGC_39"				
	/notes="Organ: pancreas; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin				

(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."

ORIGIN

```

Query Match      39.4%; Score 513.6; DB 10; Length 787;
Best Local Similarity 96.2%; Pred. No. 4.7e-115;
Matches 558; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
Db 210 GGATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 269
QY 63 TATCCCCATGGAGACCTTCAGAAA-GGTGGGGATCCCATCATCATATAGCACTACTGAGCC 121
Db 270 TATCCCCATGGAGACCTTCAGAAAATGGTGGGATCCCATCATCATATAGCACTACTGAGCC 329
QY 122 TGGCAGATATCATATTGTGGTGTCTCATCAAGTGTATTCTGGATAAATACTACTTCC 181
Db 330 TGGCAGATATCATATTGTGGTGTCTCATCAAGTGTATTCTGGATAAATACTACTTCC 389
QY 182 TCTGGGGCAGCCTCTCCACTTCATCCGAGAGAGCAGCTGTGTACGGAGAGCTGGACT 241
Db 390 TCTGGGGCAGCCTCTCCACTTCATCCGAGAGAGCAGCTGTGTACGGAGAGCTGGACT 449
QY 242 GTCCCTTGGGGAGAGCAGAGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTGCAGTGG 301
Db 450 GTCCCTTGGGGAGAGCAGAGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTGCAGTGG 509
QY 302 CAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGGCTGTGACTCGGCCACAGGAACT 361
Db 510 CAGTCCGCTCTCCAAAGACCGATCCACACTGCGAGGCTGTGACTCGGCCACAGGAACT 569
QY 362 GGTCTCTGCTGTGTTTCGACAACTTCACAGAGCTCTCGTGAGACAGCTGTAGGCAGA 421
Db 570 GGTCTCTGCTGTGTTTCGACAACTTCACAGAGCTCTCGTGAGACAGCTGTAGGCAGA 629
QY 422 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCAAGAGATCTGG 481
Db 630 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCAAGAGATCTGG 689
QY 482 ATGTTGTTGAATCAGAAAACAGCCAGAGCTTCGCATGGGNACTCAAGTGGGCCCT 541
Db 690 ATGTCGACGACATCAGAAAACAGCCAGAGCTTCGCCTTCGGACTCAAGTGGG-CCT 748
QY 542 GTCTCTCAGGCTCCCTGTCTCCCTGCACCTGTCTTGCCTG 581
Db 749 GTCTCTCAGGCTCCCTGGGTCT-CCTGCACTGTCTTGACTG 787

```

Search completed: June 8, 2004, 02:20:27
Job time : 3787 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 23:24:49 ; Search time 113 Seconds
(without alignments)
6408.946 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305

Sequence: 1 atgatactacacagtataa.....atgtctggaaggtgagctg 1305

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCURS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303.4	99.9	2038	3	US-09-008-271A-18
2	1303	99.8	2081	4	US-09-851-588-7
3	1292	99.0	2079	4	US-09-656-002-1
4	1292	99.0	2079	4	US-09-851-588-5
5	1255	11.9	2413	3	US-09-518-046-1
6	150	11.5	1479	3	US-09-342-749-1
7	150	11.5	1479	4	US-09-691-840-1
8	147	11.3	2479	3	US-09-342-749-29
9	147	11.3	2479	4	US-09-691-840-29
10	147	11.3	2479	4	US-09-685-166A-894
11	142.8	10.9	1077	3	US-08-807-151-2
12	142.8	10.9	1077	4	US-09-478-957-2
13	133.6	10.2	2416	3	US-09-261-416-1
14	132.8	9.4	2544	3	US-09-518-046-3
15	119.4	9.1	942	4	US-09-636-382A-3
16	119.2	9.1	1154	4	US-09-636-382A-1
17	112	8.6	1110	4	US-09-386-653A-1
18	112	8.6	1212	4	US-09-620-312D-431
19	111.6	8.6	696	1	US-08-508-448C-24
20	111.6	8.6	1460	4	US-09-370-838-80
21	111.6	8.6	1517	1	US-08-508-448C-15
22	111.6	8.6	2790	4	US-09-370-838-79
23	111	8.5	980	1	US-08-023-942A-30
24	110.2	8.4	901	1	US-08-508-448C-9
25	108.4	8.3	1462	1	US-09-370-838-55
26	106.4	8.2	1378	4	US-09-907-794A-262
27	106.4	8.2	1378	4	US-09-905-125A-262

28 106.4 8.2 1378 4 US-09-902-775A-262
29 106.4 8.2 1430 4 US-09-386-629-1
30 99.8 7.6 1440 1 US-07-882-202A-3
31 99.8 7.6 1440 1 US-08-021-615A-3
32 99.8 7.6 1440 1 US-08-321-777-3
33 99.8 7.6 1440 3 US-09-009-217-13
34 99.8 7.6 1440 3 US-09-009-656-13
35 99.8 7.6 1440 5 PCT-US93-04493-3
36 99.8 7.6 2462 2 US-08-479-733A-25
37 99.8 7.6 2462 3 US-08-487-427-25
38 99.8 7.6 2462 3 US-08-479-727A-25
39 99.8 7.6 2462 3 US-08-482-369A-25
40 99.8 7.6 2462 5 PCT-US95-07439-25
41 98.2 7.5 2422 1 US-08-475-845-1
42 98.2 7.5 2422 1 US-08-327-690-1
43 98.2 7.5 2422 2 US-08-660-289-1
44 98.2 7.5 2422 2 US-08-537-807-1
45 98.2 7.5 2422 2 US-08-871-003-1

ALIGNMENTS

RESULT 1

; Sequence 18, Application US/09008271A
; Patent No. 6203979

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Tom Y.

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/008,271A

; FILING DATE: 16-Jan-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: <Unknown>

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNOT13

; CLONE: 1337018

; SEQUENCE DESCRIPTION: SEQ ID NO: 18:

; US-09-008-271A-18

```
Query Match          99.9%; Score 1303.4; DB 3; Length 2081;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGATCCTGACAGTATCACTCTGAAACAGCCTCGATGTCAAAACCCCTGCGCAAAACC 60
DB 200 ATGGATCCTGACAGTATCACTCTGAAACAGCCTCGATGTCAAAACCCCTGCGCAAAACC 259
QY 61 CGTATCCCATGAGACCTTCAAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CGTATCCCATGAGACCTTCAAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 319
QY 121 TGGCGAGTATCATCATTTGGTGTGCTCATCAAGAGTATCTCGATAAATACTACTTC 180
DB 320 TGGCGAGTATCATCATTTGGTGTGCTCATCAAGAGTATCTCGATAAATACTACTTC 379
QY 181 CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGCGAGAGCTGGAC 240
DB 380 CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGCGAGAGCTGGAC 439
QY 241 TGTCTCTTGGGGAGGACGAGAGACATGTGTCAAGAGCTTCCCGAAGGCCCTGCGAGTG 300
DB 440 TGTCTCTTGGGGAGGACGAGAGACATGTGTCAAGAGCTTCCCGAAGGCCCTGCGAGTG 499
QY 301 CGAGTCCGCTCTCCAAGGACCGATCCACACTGCGAGGTGCTGGACTCGGCCACAGGGAAC 360
DB 500 CGAGTCCGCTCTCCAAGGACCGATCCACACTGCGAGGTGCTGGACTCGGCCACAGGGAAC 559
QY 361 TGGTCTCTGCTGTTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGCAG 420
DB 560 TGGTCTCTGCTGTTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGCAG 619
QY 421 ATGGGCTACAGCAGAAACCCACTTTCAGAGCTGTGGAGATGGCCAGACAGAGATCTG 480
DB 620 ATGGGCTACAGCAGAAACCCACTTTCAGAGCTGTGGAGATGGCCAGACAGAGATCTG 679
QY 481 GATGTGTTGTAATCACAGAAACAGCAGAGCTTCCGATCGCGAACTCAAGTGGGCC 540
DB 680 GATGTGTTGTAATCACAGAAACAGCAGAGCTTCCGATCGCGAACTCAAGTGGGCC 739
QY 541 TGTCTCTCAGGCTCCCTGCTCTCCCTGACATGCTGTGCTGTGGAGAGAGCTGAAAGC 600
DB 740 TGTCTCTCAGGCTCCCTGCTCTCCCTGACATGCTGTGCTGTGGAGAGAGCTGAAAGC 799
QY 601 CCCGCTGTGTGGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCGAGTCAAGTCAAGT 660
DB 800 CCCGCTGTGTGGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCGAGTCAAGTCAAGT 859
QY 661 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCTGCTGCTCAAG 720
DB 860 CAGTACGACAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCTGCTGCTCAAG 919
QY 721 CGAGCCCACTCTTTCAGGAAACATACCGATGTTCCTCACTGAGAGGTGCGGCGAGGTCA 780
DB 920 CGAGCCCACTCTTTCAGGAAACATACCGATGTTCCTCACTGAGAGGTGCGGCGAGGTCA 979
QY 781 GACAAACTGGGAGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 980 GACAAACTGGGAGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039
QY 841 CCCATGTACCCCAAGACATGACATGCGCTCATGAGAGCTGCGATTCCTCACTCACTTC 900
DB 1040 CCCATGTACCCCAAGACATGACATGCGCTCATGAGAGCTGCGATTCCTCACTCACTTC 1099
QY 901 TCAGGCAAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 1100 TCAGGCAAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1159
QY 961 CCACCTCTGGATCATTTGGAGCTTTTACGAAGCAGAGATGAGGAGAGATGCTTGACATA 1020
DB 1160 CCACCTCTGGATCATTTGGAGCTTTTACGAAGCAGAGATGAGGAGAGATGCTTGACATA 1219

RESULT 2
US-09-851-588-7
; Sequence 7, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (215)..(1528)
; OTHER INFORMATION:
US-09-851-588-7

Query Match          99.8%; Score 1303; DB 4; Length 2081;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAACCCCTGCGCAAAACCCCG 62
DB 223 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAACCCCTGCGCAAAACCCCG 282
QY 63 TATCCCATGAGACCTTCAGAAAGTGGGATTCCTCATCATAGCTATCATAGCTACTGAGCCT 122
DB 283 TATCCCATGAGACCTTCAGAAAGTGGGATTCCTCATCATAGCTACTCATAGCTACTGAGCCT 342
QY 123 GCGAGTATCATCATTTGGTGTGCTCTCATCAAGGTGATTCCTGGAATAATACTACTTCCT 182
DB 343 GCGAGTATCATCATTTGGTGTGCTCTCATCAAGGTGATTCCTGGAATAATACTACTTCCT 402
QY 183 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGAGCTG 242
DB 403 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGAGCTG 462
QY 243 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGGCTGCGAGTGC 302
DB 463 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGGCTGCGAGTGC 522
```


Db 876 AGTACGACAAACAGACAGCTGTGTGAGGAGGATCTCTGGACCCCACTGGGTCTCACGG 935
Qy 722 CAGCCCACTGTTTCAGGAAACATACCGATGTGTCACTGGAAGGTGGGGCAGGCTCAG 781
Db 936 CAGCCCACTGTTTCAGGAAACATACCGATGTGTCACTGGAAGGTGGGGCAGGCTCAG 995
Qy 782 ACMAACTGGGAGCTTCCATCCCTGGCTGTGGCCACAGATCATCATTTGAATTCACCC 841
Db 936 ACMAACTGGGAGCTTCCATCCCTGGCTGTGGCCACAGATCATCATTTGAATTCACCC 1055
Qy 842 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCTTCT 901
Db 1056 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCTTCT 1115
Qy 902 CAGGACAGTACAGCCCACTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
Db 1116 CAGGACAGTACAGCCCACTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
Qy 962 CACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAAGGAGATGTCTGACATAC 1021
Db 1176 CACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAAGGAGATGTCTGACATAC 1235
Qy 1022 TGCTCAGGCTCAGTCCAGTCAATGACAGACACAGTGCATGCAATGCAAGTCCGTACC 1081
Db 1236 TGCTCAGGCTCAGTCCAGTCAATGACAGACACAGTGCATGCAATGCAAGTCCGTACC 1295
Qy 1082 AGGGGGAAGTCAACGAGAAAGATGTGTGAGGAGTCCCGAAGGGGTGTGACACCT 1141
Db 1236 AGGGGGAAGTCAACGAGAAAGATGTGTGAGGAGTCCCGAAGGGGTGTGACACCT 1355
Qy 1142 GCCAGGGTACAGTGGTGGGCTTGTATGATACCAATCTGACAGTGGCAATGTGTGGCA 1201
Db 1356 GCCAGGGTACAGTGGTGGGCTTGTATGATACCAATCTGACAGTGGCAATGTGTGGCA 1415
Qy 1202 TGTAGTGGGCTATGGCTCGGGGCGCGAGCACCCAGGAGTATACACCAAGTCT 1261
Db 1416 TGTAGTGGGCTATGGCTCGGGGCGCGAGCACCCAGGAGTATACACCAAGTCT 1475
Qy 1262 CAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
Db 1476 CAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1519

RESULT 4

US-09-851-588-5
; Sequence 5, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-851-588-5

Query Match 99.0%; Score 1292; DB 4; Length 2079;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 3 GGATCTCTGACAGTGTCAACCTCTGAACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 62
Db 216 GGATCTCTGACAGTGTCAACCTCTGAACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 275
Qy 63 TATCCCATGAGAGACCTTCAGAAAG- GTGGGATGCCATCATCATAGCACTACTGAGCC 121
Db 276 TATCCCATGAGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
Qy 122 TGGCAGTATCATCATGTGTGTCTCATCAAGTGATTTCTGGATTAATTAATTAATTTCC 181
Db 336 TGGCAGTATCATCATGTGTGTCTCATCAAGTGATTTCTGGATTAATTAATTAATTTCC 395
Qy 182 TGTGGGGCAGCTCTCCATTTTCATCCGAGAAAGAGCTGTGTGACGAGAGCTTGGACT 241
Db 396 TGTGGGGCAGCTCTCCATTTTCATCCGAGAAAGAGCTGTGTGACGAGAGCTTGGACT 455
Qy 242 GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTGG 301
Db 456 GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTGG 515
Qy 302 CAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGAACT 361
Db 516 CAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGAACT 575
Qy 362 GGTCTCTGCTGTTCGACAACTTTCAGAGAGCTTCTGCTGAGACAGCTGTAGGACAG 421
Db 576 GGTCTCTGCTGTTCGACAACTTTCAGAGAGCTTCTGCTGAGACAGCTGTAGGACAG 635
Qy 422 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGTGAGATTTGGCCACAGCAATCTGG 481
Db 636 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGTGAGATTTGGCCACAGCAATCTGG 695
Qy 482 ATGTTTGAATTCAGAAAACAGCAGAGAGCTTTCGATGCGGAACTCAAGTGGGCTCT 541
Db 696 ATGTTTGAATTCAGAAAACAGCAGAGAGCTTTCGATGCGGAACTCAAGTGGGCTCT 755
Qy 542 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGCGGAGAGCTTGAAGAGCC 601
Db 756 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGCGGAGAGCTTGAAGAGCC 815
Qy 602 CCGTGTGGTGGGAGGAGGCTCTGTGGATTTTGGCTTTGGCAGTTCAGAGTTCAGATCC 661
Db 816 CCGTGTGGTGGGAGGAGGCTCTGTGGATTTTGGCTTTGGCAGTTCAGAGTTCAGATCC 875
Qy 662 AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGACCCCACTGGGTCTCACGG 721
Db 876 AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGACCCCACTGGGTCTCACGG 935
Qy 722 CAGCCCACTGTCTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCAG 781
Db 936 CAGCCCACTGTCTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCAG 995
Qy 782 ACAAATGGGAGCTTCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCACCC 841
Db 996 ACAAATGGGAGCTTCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCACCC 1055
Qy 842 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCTTCT 901
Db 1056 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCCTCACTCTTCT 1115
Qy 902 CAGGCACTGAGGCCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
Db 1116 CAGGCACTGAGGCCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
Qy 962 CACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAAGGAGATGTCTGACATAC 1021
Db 1176 CACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAAGGAGATGTCTGACATAC 1235
Qy 1022 TGCTCAGGCTCAGTCCAGGTCATTTGACAGACACAGTGCATGCAATGCAAGTCCGTACC 1081
Db 1236 TGCTCAGGCTCAGTCCAGGTCATTTGACAGACACAGTGCATGCAATGCAAGTCCGTACC 1295
Qy 1082 AGGGGGAAGTCAACGAGAGATGTGTGTGAGGAGTCCCGAAGGGGTGTGACACCT 1141

Db 1296 AGGGGAAAGTACCGAGAGAGATGATGTGTGAGGATCCCGGAGGGGGTGTGACACCT 1355
QY 1142 GCCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCACTGGCATCTGTGGGCA 1201
Db 1356 GCCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCACTGGCATGTGTGGGCA 1415
QY 1202 TCGTTAGTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1261
Db 1416 TCGTTAGTGGGGCTATGGCTGCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1475
QY 1262 CAGCCTATCTCACTGATCTACCAATCTGCTGGAAGGCTGAGCTG 1305
Db 1476 CAGCCTATCTCACTGATCTACCAATCTGCTGGAAGGCTGAGCTG 1519
RESULT 5
US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1
Query Match 11.9%; Score 155; DB 3; Length 2413;
Best Local Similarity 53.5%; Pred. No. 1.3e-33;
Matches 420; Conservative 0; Mismatches 350; Indels 15; Gaps 4;
QY 532 AGTGGGCCCCCTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGCTGTGGGAAGC 591
Db 717 AGGAGGGATGTGCTCTGGCCACGTTGTTACCTTGAGTGACAGCCTGTGTGTCATAGA 776
QY 592 CTG---AAGACCCCGTGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGCCCTTG 648
Db 777 AGGGGCTACAGCTCAGCATCGTGGTGGAAACATGTCTTCTCGCAGTGGCCCTGG 836
QY 649 CAGGTCAGCATCCAGTACGACAAACAGACAGTCTGTGGAGGAGCATCTGGACCCAC 708
Db 837 CAGGCCAGCTTCAGTTCCAGGGTACACCTGTGGGGGGCTCTGTATCAGCCGCTG 896
QY 709 TGGTCTCTCAGGGAGGCCACTGTT-----CAGGAAACATACCGATGTGTCAACTGG 762
Db 897 TGGATCATCACTGTGCACATGTGTTTATGACTGTGTACCTCCCAAGTCATGGACCATC 956
QY 763 AAGGTGGGGAGGCTCAGACAAACTGGGAGCTTCCCATCCCTGGCTGGCCAGATC 822
Db 957 CAGGTGGGTCTAGTTTCCCTTTTGGACAATCCAGCCCATCCCACTTTGGTGGAGAAGATT 1016
QY 823 ATCATCATTTGAATTCACACCCCATGTACCCCAAGACATGACATCGCCCTCATGAAGCTG 882
Db 1017 GTCTACACAGCATGACAGCCAAAGAGCTGGGCAATGACATCGCCCTTATGAGCTG 1076
QY 883 CAGTCCGACCTCTTCTCAGGCACAGTACGGCCCATCTGTGTGCCCTCTTTGATGAG 942
Db 1077 GCCGGGCACTCAGTTCAATGAATGATCCAGGCTGTGTGCCCTGCCCCAACTCTGAAGAG 1136
QY 943 GAGCTCACTCCAGGCCACCCCACTCTGGATCATTTGATGGGGCTTACGAGCAGAAATGGA 1002

Db 1137 AACTTCCCGCATGGAAAGTGTGCTGGAGCTCAGGATGGG---GGCCACAGAGGATGGA 1193
QY 1003 GGGAAGATGTCTGACATATCTGCTGAGGGCTGAGTCCAGGTCATTCAGACAGCACGGTGC 1062
Db 1194 GGTGACGCTCCCTGCTGTAACACAGCGGCGCTCCCTTTGATTTCCAAAGATCTCTG 1253
QY 1063 AATGCAAGCATGCGTACCAAGGGGAAAGTCAACGAGAAAGATGATGTGCAAGCATCCCG 1122
Db 1254 AACCAAGGAGCGTGTACGGTGGCATCATCTCCCCCTCCATGCTCTGCGCGGCTACCTG 1313
QY 1123 GAAGGGGGTGTGGACACTGCCAGGGTGACAGTGGTGGGGCCCTGATGTACCAA---TCT 1179
Db 1314 ACGGGTGGCGTGAACAGCTGCCAGGGGGACAGCGGGGGCCCTGCTGTGTCAAGAGAG 1373
QY 1180 GACCACTGGCATGTGCTGGGCTCGTTAGCTGGGGCTATGCTGGGGGGCCCGAGCACC 1239
Db 1374 AGGCTGTGGAAGTTAGTGGAGGACCAAGCTTTGGATCGGCTGCGCAGAGGTGAACAAG 1433
QY 1240 CCAGGAGTATACCAAGGCTCTCAGCCTATCTCACTGGATCTACAACTCTTGGAAAGGCT 1299
Db 1434 CTTGGGGTGTACACCGCTGTCACTCTTCTGGACTGGATCCAGCAGCAGATGGAGAGA 1493
QY 1300 GAGCT 1304
Db 1494 GACCT 1498
RESULT 6
US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1466)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1471)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
; FEATURE:

NAME/KEY: allele
 LOCATION: (478)
 OTHER INFORMATION: This base can be G or A with G being the more
 OTHER INFORMATION: common allele. The codon will change from Val to
 OTHER INFORMATION: Met.
 FEATURE:
 NAME/KEY: allele
 LOCATION: (777)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. The codon is unaffected with both
 OTHER INFORMATION: alleles encoding Gly.
 FEATURE:
 NAME/KEY: allele
 LOCATION: (768)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. This is a silent polymorphism.
 FEATURE:
 NAME/KEY: allele
 LOCATION: (834)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. This is a silent polymorphism.
 FEATURE:
 NAME/KEY: allele
 LOCATION: (625)
 OTHER INFORMATION: This base can be T or A with T being the more
 OTHER INFORMATION: common allele. The codon will change from Phe to
 OTHER INFORMATION: Ile
 US-09-342-749-1

Query Match 11.5%; Score 150; DB 3; Length 1479;
 Best Local Similarity 51.7%; Pred. No. 2.6e-32;
 Matches 534; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCTCTCCAGGACCGATCCACATCTGAGGTGCTGGACTCGGCCACAGGAACTGG 363
 DB 445 GTTCGCTCTACGGACCAAACTTCATCTCTCAGGTGTAATCTATCTCAGAGAACTCTGG 504
 QY 364 TTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGACAGAGCTGTGAGCAGATG 423
 DB 505 CACCCTGTGTGCAAGACGACTGGAACAGAACTACGGCGGGCGCTGAGGAGCATG 564
 QY 424 GCGTACAGAGAAACCCATCTTCAGAGCTGTGAGATGTCGCCAGACAGATCTGGAT 483
 DB 565 GCGTATAAGAAATAATTTTACTCTAG-CAAAGGAATAGTGATGACAGCGGATCCACG 623
 QY 484 GTTGTGAAATCACAGAAACAGCGAGCTTCG-----CATCGGAACTCAAG 533
 DB 624 CTTTATGAACCTGAACAAGTCCGGCAATGTCGATATCTATAAAACTGTACCACAG 683
 QY 534 TGGGCGCTGTCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
 DB 684 TGATGCTGCTTCTTCAAAAGCAGTGGTCTTCTTACGCTGTATAGCTCGGGGTCAAAT 743
 QY 594 GAAGACCCCTC-----GTGCTGCTGGGAGGAGCGCTCTGTGATTTCTTGCC 644
 DB 744 GAATCAAGCCCGCAGAGAGATCTGGCGGCGAGAGCGGCTCCCGGGGCTGGCC 803
 QY 645 TTGGCAGCTCAGCATCCAGTACGACAAACAGACGCTCTGTGGAGGAGCATCTGGACCC 704
 DB 804 CTGGCAGGTCAGCTGACCTCCAGACGCTCCACGCTGTGGAGGCTCCATCATCACCCC 863
 QY 705 CCACTGGGCTCTCAGGAGCCCTGCTTCAGGAACAT-----ACCGATGTGTCA 757
 DB 864 CAGTGTGATCTGTGAGCGCGCCCTCTGGTGGAACAACTCTTAACATCATGGCATG 923
 QY 758 ACTGGAAGTGTGGGCGGCTCAGACAAACTGGGAGCTTCCATCCCTGCTGTGGCCA 817
 DB 924 GACGCAATTTGCGGATTTTTCAGACAACTTTTCATGTTCTATGAGCGGATACCAAGT 983
 QY 818 AGATCATCATCTGATTAATCAACCCCATGT-----ACCCCAAGACATGACATCGCCT 872
 DB 984 AGAAAAAGTGATTTCTCATCTCAAAATATAGTCCAGACCAAGAAACAATGATTCGCT 1043

QY 873 CATGAAGCTGCAGTTCCCACTCACCTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTT 932
 DB 1044 GATGAAGCTGCAGAGCCCTCTGACCTTTCAACGACCTAGTGAACACAGTGTGTCTGCCAA 1103
 QY 933 CTTTGTATGAGGAGCTCACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAA 992
 DB 1104 CCAGGCATGATGCTGCAGCCAGAACAGCTCTGCTGGATTTCCGGGTGGGGGCCACCGA 1163
 QY 993 GCAGATGAGGAGAGATGCTGTGACATATGCTGTGAGGCGTCAGTCCAGGTCAATGACAG 1052
 DB 1164 GGAGAA---AGGGAAGACCTCAGAAGTGTCTGAACTGCAAGGCTGCTCTCATTTGAGAC 1220
 QY 1053 CACACGGTGTCAATGACAGCATGCGTACCGAGGGGAAAGTCACCGAGAAAGATGATGTGTC 1112
 DB 1221 ACAGAGATCAACAGCAGATATGCTATGACAACTGATCACCAGCCATGATCTGTGC 1280
 QY 1113 AGCATCCCGGAAGGGGTGTGACACCTGCCAGGGTGACAGTGTGGCCCTG---AT 1169
 DB 1281 CGGCTTCTCTCAGGGGAACGTCGATTTCTTCCAGGGTGACAGTGGAGGCTCTCTGTCTAC 1340
 QY 1170 GTACCAATCTGACCAAGTGGCATGTGTGGGCATCGTTAGCTGGGCTATGGCTGGGGGG 1229
 DB 1341 TTCGAAGAACAATATCTGTGTGCTGATAGGGGATCAAGCTGGGTTCTGGCTGTGCCAA 1400
 QY 1230 CCGAGACCCCGAGGAGTATACACCAAGGTCTCAGCTTATCTCAACTGATCTACAATGT 1289
 DB 1401 AGCTTACAGACAGGAGTGTACGGGAATGTGATGTATTACGGAGCTGGATTTATCGACA 1460
 QY 1290 CTGGAAGGCTGA 1301
 DB 1461 AATGAGGCGAGA 1472

RESULT 7

US-09-691-840-1
 ; Sequence 1, Application US/09691840
 ; Patent No. 6444419
 ; GENERAL INFORMATION:
 ; APPLICANT: Wong, Alexander K.C.
 ; APPLICANT: Tavtigian, Sean V.
 ; APPLICANT: Teng, David H.-F.
 ; APPLICANT: Myriad Genetics, Inc.
 ; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
 ; FILE REFERENCE: 2318-202
 ; CURRENT APPLICATION NUMBER: US/09/691,840
 ; CURRENT FILING DATE: 2000-10-18
 ; PRIOR APPLICATION NUMBER: US/09/342,749
 ; PRIOR FILING DATE: 1999-06-29
 ; PRIOR APPLICATION NUMBER: US 60/091,044
 ; PRIOR FILING DATE: 1998-06-29
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1479
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1476)
 ; NAME/KEY: conflict
 ; LOCATION: (724)
 ; OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
 ; NAME/KEY: conflict
 ; LOCATION: (985)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1347)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1466)
 ; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1471)

OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to Ile
US-09-691-840-1

Query Match 11.5%; Score 150; DB 4; Length 1479;
Best Local Similarity 51.7%; Pred. No. 2.6e-32;
Matches 534; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 363
DB 445 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 504
QY 364 TTCCTGCTGTTTCGACAACTTACAGAACTCTCGCTGAGACAGCTGTAGGACAGATG 423
DB 505 CACCTGTGTCCAGACGACTGGAACAGAACTACCGGGCGGGCTGTGAGGACATG 564
QY 424 GGCTACAGCAGCAACACCTTTTCAGAGCTGTGAGATTGGCCAGACAGGATCTGGAT 483
DB 565 GGCTATAGATATTTTACTCTAG-CCAGGATAGTGTGATGACAGCGATCCACAG 623
QY 484 GTTGTGAAATACAGAAAAAGCAGGAGTTGG------CATGGGAATCAAG 533
DB 624 CTTTATGAACTGAACACAGTCCGGAATGTGCATATCTATAAAAACTGTACCACAG 683
QY 534 TGGGCTCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
DB 684 TGATGCTGTCTTCAAAGCAGTGGTTCTTTAGCTGTATAGCTGCGGGGTCACTT 743
QY 594 GAAGACCCCTCC------GTGTGGTGGTGGGAGGAGGCTCTGTGGATTTCTGGCC 644
DB 744 GAACCTCAAGCCGACAGACGATGCTGGCGGAGAGCGCTCCCGGGGCTGGCC 803
QY 645 TTGGCAGTTCAGCTCAGTACAGAAACAGCAGCAGTCTGTGGGAGGACATCTGGACCC 704
DB 804 CTGGCAGTTCAGCTCAGTCCAGAACGTCACGTTGCGGAGGCTCCATCATCACCCC 863
QY 705 CCATGCTGCTCAGCGAGCCCTGCTTCCAGGAACAT------ACCGATGTGTCA 757
DB 864 CGATGGATCTGACAGCGCCCTGCTGGTGAACAACTCTTAACATCCATGCAATG 923
QY 758 ACTGGAAGTGGGCGAGGCTCAGACAACTGGGAGCTTCCCATCCCTGCTGTGGCCA 817
DB 924 GACGGCATTTGCGGGAATTTGAGACAATCTTATGTTCTATGGAGCCGATACCAAGT 983
QY 818 AGATCATCATATTGAATTAACCCCATGT------ACCCCAAGACATGATCGCCCT 872
DB 984 AGAAAAAGTATTTCTCATCAATTAATGACTCCAGAACCAAGAACATGATGCT 1043
QY 873 CATGAAGTGCAGTTCCCACTCATTCTCAGGCACAGTCAAGGCCCATCTCTGCTGCCCT 932
DB 1044 GATGAAGTGCAGAGGCTCTGACTTTCAACGACCTAGTGAACAGGCTGTCTGTGCCAA 1103

QY 933 CTTTGTAGAGGAGCTCATTCCAGCCACCCACTCTGGATCATTTGGATGGGGCTTTACGAA 992
DB 1104 CCAGGCATGATCTCGAGCCAGAACAGCTCTGCTGATTTCCGGTGGGGGCCACCGA 1163
QY 993 GCAGATGAGGAGGAGATCTGTGACATATCTGTGAGGGCTCAGTCCAGCTCATTCACAG 1052
DB 1164 GGAGAA---AGGGAAGCTCTCAGAGTGTGAAACGCTGCAAGTCTTCTCATTTAGAC 1220
QY 1053 CACACGCTCAATGACAGCAATGCTGACAGGGGAAAGTCAACGAGAAAGATGATGTGTC 1112
DB 1221 ACAGAGATCAACACAGCAGATATGTCTATGACAACTGTATGACACACAGCCATGATCTGTGC 1280
QY 1113 AGGCATCCCGAAGGGGTGTGGACACCTCCAGGGTGCAGTGTGGGGCCCTG---AT 1169
DB 1281 CGGCTTCTCGAGGGGAAGCTGATTTCTTGCCAGGGTGACAGTGGAGGCTCTGTGTCAC 1340
QY 1170 GTACCAATCTGACCACTGCTGCTGGGCATCGTTAGCTGGGGCTATGGCTGCGGGGG 1229
DB 1341 TTGGAAGACAATATCTGGTGGCTGATAGGGGATACAGCTGGGGTTCTGGCTGTGCCAA 1400
QY 1230 CCGGACACCCAGGAGTATACACAAAGTCTCAGCCTATCTCACTGGATCTACATGT 1289
DB 1401 AGCTTACAGACCAAGAGTGTACGGGAATGTGATGGTATTACCGGACTGGATTATCGACA 1460
QY 1290 CTGGAAGGCTGA 1301
DB 1461 AATGAGGGCAGA 1472

RESULT 8
US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; EARLIER FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-749-29

Query Match 11.3%; Score 147; DB 3; Length 2479;
Best Local Similarity 51.8%; Pred. No. 2.3e-31;
Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;
QY 304 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 363
DB 501 GTTCGCTCTACGGACCAAACTTCATCTTCAGATGTACTCATCTCAGAGGAAGTCTCTGG 560
QY 364 TTCTCTGCTGTTTCGACAACTTCAAGAGCTCTCGCTGAGACAGCTGTAGGACAGATG 423
DB 561 CACCTCTGTGTCCAGACGACTGGAACAGAACTACGGGCGGGCTGTGAGGACATG 620
QY 424 GGCTACAGCAGCAACCCACTTTTCAGAGTGTGGAGATTGGCCACAGCAGGATCTGGAT 483
DB 621 GGCTATAGATATTTTACTCTAG-CCAAGGAATAGTGGATGACAGCGGATCCACAG 679
QY 484 GTTGTGAAATCAGAAAAAGCAGGAGCTTCG------CATGGGAATCAAG 533
DB 680 CTTTATGAACTGAACACAGTGTGATGATGCTGATATCTATAAAAACTGTACACAG 739
QY 534 TGGGCTCTGCTCTCAGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593

740 TGATGCTGTTCTTCAAAAGCAGTGGTTCTTTAGCGTGTGTAGCGCTGCGGGTCAACTT 799
594 GAAGACCCCGCC-----GTGCGTGGGTGGGAGGAGGCGCTCTGTGGATTCTTGCC 644
800 GACTCAAGCCCGCAGACAGAGATCGTGGCGGTGAGAGCGCGCTCCCGGGGCGCTGGCC 859
645 TTGGCAGGTGACATCCAGTACGACAAACAGACAGTCTGTGTGGAGGAGCATCTTGGACC 704
860 CTGGCAGGTGACCTGACGCTCCAGACAGTCCAGCTGTGGAGGCTCCATCATCACCCC 919
705 CCACTGGGTCTCAGCGAGCCCACTGTCTCAGGAACAT-----ACCGATGTGTCA 757
920 CGAGTGGATCGTGAAGCGCCCACTGTGTGGAACAACTCTTACAACTCCATGGCATG 979
758 ACTGGAAGGTGCGGCGAGCTCAGACAAACTGGGAGCTTCCCATCCCTGCTGTGGCCA 817
980 GACGCGATTGCGGGGATTTTGAACAATCTTTATGTTCTATGAGCGGATACCAAGT 1039
818 AGATCATCATCAATTGAATTCACCCATGTACC-----CCAAAGACATGATGCGCT 872
1040 ACAAAGATGATTTCTCATCCAAATATGACTCCAAAGACCAAGAACAAATGATGATG 1099
873 CATGAAGTGTGAGTTCCTCACTTCTCAGGACAGTCCAGGCGCATCTGTGCGCCTT 932
1100 GATGAAGCTGCAAGAGCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTGCCAA 1159
933 CTTTATGAGGAGTCACTCCAGCCACCCCTCTGAGTCTGTGGAGGAGCATCTTGACCC 992
1160 CCCAGGATGATGCTGACAGCAAGCTCTGCTGGATTTCCGGGTGGGGGCGCCACCGA 1219
993 GCAGATGAGGAGGAGTGTCTGACATCTGTGAGGCGTCTGAGGCGTCTGAGGCGTCTGACAG 1052
1220 GGAGAA---AGGAGAGCTCAGAGTGTGAACTGCTGAACTGCTTCTCATTTGAGAC 1276
1053 CACCGGTGCAATGACAGATGCTGACCAAGGAGTCCAGGAGGAGTCCAGGAGATGATGTGTC 1112
1277 ACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACCAGGCGATGATCTGTGC 1336
1113 AGGCATCCGGAGGGGTGTGGACACCTGCGAGGTCACAGTGTGTGGCGCCCTG---AT 1169
1337 CGGCTTCTGAGGGGAACTGATTTCTGCGAGGTCAGATGAGGCGCTCTGCTCAC 1396
1170 GTACCAATCTGACAGTGGCATGTGGTGGCATCGTTAGCTGGGCTATGCTCGGGGG 1229
1397 TTCGAACAACAATATCTGGTGTGATAGGGGATACAAAGTGGGTTCTGGCTGTGCCAA 1456
1230 CCGGAGCACCAGAGTATACCAAGGCTCTCAGCTATCTCACTGGATCTACAATGT 1289
1457 AGCTTACAGACCAGGAGTGTACGGGAATGTGATGTATTCACGGACTGGATTTATCGACA 1516
1290 CTGGAAGGC 1298
1517 AATGAGGC 1525

RESULT 9

US-09-691-840-29
; Sequence 29, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavcigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TSPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-691-840-29

Query Match 11.3%; Score 147; DB 4; Length 2479;
Best Local Similarity 51.6%; Pred. No. 2.3e-31;
Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCGCGCTCTCAAGGACCGATCCACACTGCAAGGTGCTGAGTCTGGCCACAGGGAATGG 363
DB 501 GTTCGGCTCTACGGACCAAACTTCATCCTTCAGATGACTCATCTCAGAGGAATCTCTGG 560
QY 364 TTCTCTGCTGTGTTGACAACTTCACAGAACTTCGTGTGACAGACCTGTAGGAGATG 423
DB 561 CACCTGTGTGCCAAGACGACTGGAACGAACTACGGCGGGCGCTGCAAGGACATG 620
QY 424 GGTACAGAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTGGAT 483
DB 621 GGCTATAAGAAATAATTTTACTCTAG-CCAAGAAATAGTGGATGACGGGATCCACAG 679
QY 484 GTTGTGAAATCACAGAAAACAGCCAGGAGCTTCG-----CATGCGGAATCTCAAG 533
DB 680 CTTTATGAAACTGAACACAGTCCGCAATGTCGATATCTATAAAACTCTACACAG 739
QY 534 TGGGCGCTGTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTGCTGTGGAAAGCT 593
DB 740 TGAATGCTGTCTTCAAAAGCAGTGGTTCTTTACGCTGTTTAGCTGCGGGGTCAACT 799
QY 594 GAAGACCCCGCC-----GTGCTGGTGGGAGGAGGCTCTGTGGATTCTTGGCC 644
DB 800 GAACTCAAGCGCCAGCAGGAGTCTGGCGGTGAGAGCGCTCCCGGGGCGCTGGCC 859
QY 645 TTGGCAGGTGACATCCAGTACGACAAACAGCAGTCTGTGGAGGAGCATCTTGACCC 704
DB 860 CTGGCAGGTGACCTGACGCTCAGAACCTGCAAGCTGTGGAGGCTCCATCATCACCCC 919
QY 705 CCACTGGTCTCAGCGAGCCCACTGTCTCAGGAACAT-----ACCGATGTGTCA 757
DB 920 CGAGTGGATCGTGAAGCGCCCACTGCTGCTGAAACAACTCTTAACTCCATGCAATG 979
QY 758 ACTGGAAGGTGCGGCGAGGCTCAGACAAACTGGGAGCTTCCCATCCCTGCTGTGGCCA 817
DB 980 GACGCGATTGCGGGGATTTTGAACAATCTTTCATGTTCTATGAGCGGATACCAAGT 1039
QY 818 AGATCATCATCAATTGAATTCACCCCATGTACC-----CCAAAGACATGATGATGCT 872
DB 1040 ACAAAGATGATTTCTCATCCAAATATGACTCCAAAGACCAAGAACAAATGATGATG 1099
QY 873 CATGAAGTGTGAGTTCCTCACTTCTCAGGACAGTCCAGGCGCATCTGTGCGCCTT 932
DB 1100 GATGAAGCTGCAAGAGCTCTGACTTTCAACGACCTAGTGAACCAAGTGTGTGCCAA 1159
QY 933 CTTTATGAGGAGTCACTCCAGCCACCCCTCTGAGTCTGTGGAGGAGCATCTTGACCC 992
DB 1160 CCCAGGATGATGCTGACAGCAAGCTCTGCTGGATTTCCGGGTGGGGGCGCCACCGA 1219
QY 993 GCAGATGAGGAGGAGTGTCTGACATCTGTGAGGCGTCTGAGGCGTCTGAGGCGTCTGACAG 1052
DB 1220 GGAGAA---AGGAGAGCTCAGAGTGTGAACTGCTGAACTGCTTCTCATTTGAGAC 1276
QY 1053 CACCGGTGCAATGACAGATGCTGACCAAGGAGTCCAGGAGGAGTCCAGGAGATGATGTGTC 1112
DB 1277 ACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACCAGGCGATGATCTGTGC 1336
QY 1113 AGGCATCCGGAGGGGTGTGGACACCTGCGAGGTCACAGTGTGTGGCGCCCTG---AT 1169
DB 1337 CGGCTTCTGAGGGGAACTGATTTCTGCGAGGTCAGATGAGGCGCTCTGCTCAC 1396
QY 1170 GTACCAATCTGACAGTGGCATGTGGTGGCATCGTTAGCTGGGCTATGCTCGGGGG 1229
DB 1397 TTCGAACAACAATATCTGGTGTGATAGGGGATACAAAGTGGGTTCTGGCTGTGCCAA 1456
QY 1230 CCGGAGCACCAGAGTATACCAAGGCTCTCAGCTATCTCACTGGATCTACAATGT 1289
DB 1457 AGCTTACAGACCAGGAGTGTACGGGAATGTGATGTATTCACGGACTGGATTTATCGACA 1516
QY 1290 CTGGAAGGC 1298
DB 1517 AATGAGGC 1525

Db 1397 TTCGAACAAATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTTGCTGTGCGAA 1456
QY 1230 CCCGAGCAGCCAGGATATACCAAGGCTCTCAGCCCTATCTCAACTGGATCTCAATGT 1289
Db 1457 AGCTTACAGACAGGAGTGTACGGGATGTGATGTTACCGGACTGGATTTATCGACA 1516
QY 1290 CTGGAGGC 1298
Db 1517 AATGAAGC 1525

RESULT 10

US-09-685-166A-894

; Sequence 894, Application US/09685166A

; Patent No. 6630305

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darriack

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C21

; CURRENT APPLICATION NUMBER: US/09/685,166A

; NUMBER OF SEQ ID NOS: 898

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 894

; LENGTH: 2479

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-685-166A-894

Query Match 11.3%; Score 147; DB 4; Length 2479;

Best Local Similarity 51.6%; Fred.No. 2.3e-31;

Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCTCTCCAGGACCGATCCACTGCGAGGTGCTGGACTCGGCCACAGGGAATGG 363
Db 501 GTTCGCTCTAGGACCAAACTTCATCCTTCAGATGTACTCATCTCAGAGGAAGTCCTGG 560
QY 364 TTCTTGCTGTTTCGACAACTTCAGAACTCTCGCTGAGACAGCTGTAGGAGATG 423
Db 561 CACCTGTGTGCAAGACGACTGGAACAGAACTACGGCGGGCGGCTGCAGGAGCATG 620
QY 424 GGCTACAGCAGCAACCCACTTTCAGACTGTGGAGATTGCCAGACAGGATCTGGAT 483
Db 621 GGCTATAGATAATTTTACTCTAG-CAAGGATAGTGATGACAGCGATCCACAG 679
QY 484 GTTGTGAAATCAGAGAAACAGCCAGGAGTTGG-----CATGCGGAATCAAG 533
Db 680 CTTTATGAACTGAACACAGAGTCCGCAATGTGATATCTATAAAAACTGTACACAG 739
QY 534 TGGGCGCTGTCTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGGAGAGCT 593
Db 740 TGATCCCTGTCTTCAAAGCAGTGGTTCTTACGCTGTTAGCCTCGGGGGTCAATT 799
QY 594 GAAAGACCCCCC-----GTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGCC 644

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORN0T01
; CLONE: 556016
; US-08-807-151-2

Query Match          10.9%; Score 142.8; DB 3; Length 1077;
Best Local Similarity 53.3%; Pred. No. 2.4e-30;
Matches 428; Conservative 0; Mismatches 348; Indels 27; Gaps 5;

QY 523 CGGAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGCCTGT 582
DB 146 CTTGACCAAGTGAAGAGTATTTCTCATCCAAATTATGACTCCAGACCAAGAAACAAT 205
QY 583 GGAAGAGCCTGAAGACCCCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGCCTGT 633
DB 206 GGGGTCAACTTGAACCTCAAGCGCCGAGCAGGATCGTGGCGGCGAGCGCGCTCCCG 265
QY 634 GATTCTTGCCCTTGGCAGTCAAGTACGACAAACAGCAGCTCTGTGGAGGAGC 693
DB 266 GGGGCTGGCCCTGCAGGTGAGCTGACCTGCAGCTCAGACGCTCCAGCTGTGGAGGCTCC 325
QY 694 ATCTGGACCCCACTGGGTCTCTACGGAGCCCACTGCTTCAAGAAACAT-AC 746
DB 326 ATCATCACCCCGAGTGGATCGTACAGCCGCCCACTGCTGGAGAAACCTCTTAACAAT 385
QY 747 CGATGTGTTCACTGGAAGTGGCGGAGGCTCAGACAACTGGCGAGCTTCCCATCCT 806
DB 386 CCATGGCATGGACGGCATTTGCGGGGATTTGAGACAACTCTTCATGTTCTATGGAGCC 445
QY 807 GGCT-----GTGGCCAAAGATCATCATTAATGAATTCACCCCATGTACCCCAAGACAAT 861
DB 446 GGATACCAAGTAGAAGAGTATTTCTCATCCAAATTATGACTCCAGACCAAGAAACAAT 505
QY 862 GACATGCGCCTCATGAAGTGCAGTTCCCACTCACTTCTCAGCAGCAGTCAAGCCCATC 921
DB 506 GACATGCGCTGATGAAGTGCAGAGGCTCTGACTTTTCAACGACCTAGTGAACACCAAGT 565
QY 922 TGCTGCGCCCTTTTGTATGAGAGGCTCACTCCAGCCACCCCACTCTGGATCATTTGGATGG 981
DB 566 TGCTGCGCCCAACCCAGGATGATGCTGCGCCAGACAGCTCTGCTGATTTCCGGGTGG 625
QY 982 GCTTTAGACAGCAAGTGGAGGAGATGTCTGACATCTGCTGAGGCGTCACTGAGTCCAG 1041
DB 626 GGGGCCACCGAGGAGAA---AGGGAAGACCTCAGAAGTGTCTGAACGCTGCGCAAGTGTCT 682
QY 1042 GTCAATTGACAGCACAGGTGCAATCCAGACGATGCTTACCGGGGGAAGTCAACGAGAAG 1101
DB 683 CTCATTGACACACAGATGACACAGCAGATATGTCTATGACACCTGATCACACAGGCC 742
QY 1102 ATGATGTGTGAGGATCCCGGAAGGGGTGTGACACCTCCAGGGGTGACAGTGTGGG 1161
DB 743 ATGATCTGTGCGCGTCTCTCAGCGGGAACGTCGATTTCTTCCAGGGGTGACAGTGGAGG 802

1162 CCCTG---ATGTACCAATCTGACAGTGGCATGTGGTGGCATCTGTAGCTGGGGCTAT 1218
DB 803 CNTCTGGTCACTTCGAAGAACAATATCTGTGGCTGATAGGGGATACAAAGCTGGGTTCT 862
QY 1219 GGCTGGGGGGCGGAGCAGCCCGAGGATATACCAAGGTCTCAGCCCTATCTCAACTGG 1278
DB 863 GGCTGTGCCAAAGCTTACAGACCAAGAGGTGACGGGAATGTGATGTTACGGACTGG 922
QY 1279 ATCTACAATGTCTGGAAGCTGA 1301
DB 923 ATTTATCGAACAATGAGGGCAGA 945

RESULT 12
US-09-478-957-2
; Sequence 2, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORN0T01
; CLONE: 556016
; US-09-478-957-2

Query Match          10.9%; Score 142.8; DB 4; Length 1077;
Best Local Similarity 53.3%; Pred. No. 2.4e-30;
Matches 428; Conservative 0; Mismatches 348; Indels 27; Gaps 5;

QY 523 CGGAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGCCTGT 582
DB 146 CTTGACCAAGTGAAGAGTATTTCTCATCCAAAGCAGTGGTTCTTTACGCTGTATAGCCTGC 205
QY 583 GGAAGAGCCTGAAGACCCCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGCCTGT 582
DB 146 CTTGACCAAGTGAAGAGTATTTCTCATCCAAAGCAGTGGTTCTTTACGCTGTATAGCCTGC 205
QY 583 GGAAGAGCCTGAAGACCCCGCTCTCTCAGGCTCCCTGGTGGGGAGGAGCCCTCTGTG 633
DB 206 GGGGTCAACTTGAACCTCAAGCGCCGAGCAGGATCGTGGCGGCGAGAGCGCGCTCCCG 265
QY 634 GATTCTTGCCCTTGGCAGTCAAGTACGACAAACAGCAGCTCTGTGGAGGAGC 693
DB 266 GGGGCTGGCCCTGCAGGTGAGCTGACCTGCAGCTCAGACGCTCCAGCTGTGGAGGCTCC 325
QY 694 ATCTGGACCCCACTGGGTCTCTACGGAGCCCACTGCTTCAAGAAACAT-AC 746
DB 326 ATCATCACCCCGAGTGGATCGTACAGCCGCCCACTGCTGGAGAAACCTCTTAACAAT 385
QY 747 CGATGTGTTCACTGGAAGTGGCGGAGGCTCAGACAACTGGCGAGCTTCCCATCCT 806
DB 386 CCATGGCATGGACGGCATTTGCGGGGATTTGAGACAACTCTTCATGTTCTATGGAGCC 445
QY 807 GGCT-----GTGGCCAAAGATCATCATTAATGAATTCACCCCATGTACCCCAAGACAAT 861
DB 446 GGATACCAAGTAGAAGAGTATTTCTCATCCAAATTATGACTCCAGACCAAGAAACAAT 505
QY 862 GACATGCGCCTCATGAAGTGCAGTTCCCACTCACTTCTCAGCAGCAGTCAAGCCCATC 921
DB 506 GACATGCGCTGATGAAGTGCAGAGGCTCTGACTTTTCAACGACCTAGTGAACACCAAGT 565
QY 922 TGCTGCGCCCTTTTGTATGAGAGGCTCACTCCAGCCACCCCACTCTGGATCATTTGGATGG 981
DB 566 TGCTGCGCCCAACCCAGGATGATGCTGCGCCAGACAGCTCTGCTGATTTCCGGGTGG 625
QY 982 GCTTTAGACAGCAAGTGGAGGAGATGTCTGACATCTGCTGAGGCGTCACTGAGTCCAG 1041
DB 626 GGGGCCACCGAGGAGAA---AGGGAAGACCTCAGAAGTGTCTGAACGCTGCGCAAGTGTCT 682
QY 1042 GTCAATTGACAGCACAGGTGCAATCCAGACGATGCTTACCGGGGGAAGTCAACGAGAAG 1101
DB 683 CTCATTGACACACAGATGACACAGCAGATATGTCTATGACACCTGATCACACAGGCC 742
QY 1102 ATGATGTGTGAGGATCCCGGAAGGGGTGTGACACCTCCAGGGGTGACAGTGTGGG 1161
DB 743 ATGATCTGTGCGCGTCTCTCAGCGGGAACGTCGATTTCTTCCAGGGGTGACAGTGGAGG 802
```

```

RESULT 13
US-09-261-416-1
; Sequence 1, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 144..1511
; OTHER INFORMATION: CDS
US-09-261-416-1
Query Match      10.2%      Score 133.6; DB 3; Length 2416;
Best Local Similarity 53.4%      Pred. No. 1.3e-27;
Matches 421; Conservative 0; Mismatches 349; Indels 18; Gaps 6;

```

SEQ ID NO 3
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: CDS
OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

Query Match 9.4%; Score 122.8; DB 3; Length 2544;
Best Local Similarity 55.3%; Pred. No. 1.5e-24;
Matches 282; Conservative 0; Mismatches 22; Indels 6; Gaps 2;
QY 798 CCATCCCTGGCTGTGGCAAGATCATCATCATGAAATTCACCCCAAGTACCCCAAGA 857
DB 1123 CCATCCCACTTGGTGGAGAGATTGTCTACACAGCAAGTACAAGCCAAAGAGGCTGGG 1182
QY 858 CAATGACATCGCCCTCATGAGCTGAGTCCCACTCACTTCTCAGGCACAGTCAGGCC 917
DB 1183 CAATGACATCGCCCTCATGAGCTGCGCGGCCACTCACGTTCAATGAATGATCCAGCC 1242
QY 918 CATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCCACTCTCTGGATCATGG 977
DB 1243 TGTGTGCTGCCCACTCTGAAGAACTTCCCGATGGAAGTGTGCTGGAAGTCAAG 1302
QY 978 ATGGGGCTTTAGCAAGCAGATGAGAGGAAGATGTCTGACATCTGCTGAGGGCTCAGT 1037
DB 1303 ATGGGG---GGCCACAGAGGATGGAGGTGACGCTCCCTCTGTAACCAAGCGGGCGGT 1359
QY 1038 CCAGGTCTATTGACACACACAGGTGCAATGACAGACGATCGTACCAGGGGGAAGTCAACGA 1097
DB 1360 CCTTTGATTTCCACAGATCTGCACACAGGAGCTGTACGGTGCATCATCTCCCC 1419
QY 1098 GAAGATGATGTGACGATCCCGAAGGGGTGTGACACCTGCCAGGTGACAGTGG 1157
DB 1420 CTCCATGCTCTCGCGGGCTACTCTGACGGTGGCGTGGACAGCTGCCAGGGGACAGCGG 1479
QY 1158 TGGGCCCCGTGATACCAATCTGAC---CAGTGGATGTGTGGGCGATCGTTAGCTGGG 1214
DB 1480 GGGGCCCCGTGTGTCTAAGAGAGGAGCTGTGGAGGTGTAGTGGAGCGACCACTTTGG 1539
QY 1215 CTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACCAAGGTCTCAGCTTATCAA 1274
DB 1540 CATCGCTGCGCAGAGGTGAACAACCTGGGTGTACACCCGTGTACACCTCTCTCTCGGA 1599
QY 1275 CTGGATCTCAATGTCTGGAAGCTGAGCT 1304
DB 1600 CTGGATCCACGACGATGGAGAGAGCT 1629

RESULT 15

US-09-636-382A-3
Sequence 3, Application US/09636382A
Patent No. 6514741
GENERAL INFORMATION:
APPLICANT: Presnell, Scott R.
APPLICANT: Taft, David W.
TITLE OF INVENTION: TRYPTASE-LIKE POLYPEPTIDE ZTRYPI
FILE REFERENCE: 99-21
CURRENT APPLICATION NUMBER: US/09/636,382A
CURRENT FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: US 60/149,563
PRIOR FILING DATE: 1999-08-18
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 942
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence of mouse ztrypi1
OTHER INFORMATION: (SEQ ID NO:2)
NAME/KEY: misc_feature

LOCATION: (1)...(942)
OTHER INFORMATION: n = A,T,C or G
US-09-636-382A-3

Query Match 9.1%; Score 119.4; DB 4; Length 942;
Best Local Similarity 33.8%; Pred. No. 8.4e-24;
Matches 247; Conservative 122; Mismatches 337; Indels 24; Gaps 4;
QY 578 CTTGTGGGAAGAGCCTGAAGACCCCGCTGTGGTGGGGAGGAGGAGGCTCTGTGGATT 637
DB 101 SNTGYGNCARACNAAYATHACNTGYAARGTNGTNAAYGNAARGCNGTNGARGINGNA 160
QY 638 CTTGGCTTGGCAGGTGACGATCCAGTACGACAAACAGCACGCTCTGTGAGGAGCATCC 697
DB 161 ARTGGCCTGGCARGTNGSNATHYNTTYYTNGNATGTAYATHGTGYWSNGNWSYNTNA 220
QY 698 TGGACCCCACTGGGTCCTCAGGCGAGCCCACTGTCTCAGGAAACATACCGATGTGTCA 757
DB 221 THCAYCAYCAYTGGATHYTNACNGCNCNCAVTGYTNCARMGNWSNARAAYCCGNA 280
QY 758 ACTGGAAGTGGGGCAGGCTCAGACAACT---GGGCGAGTTCCTCCATCCCTGGCTGG 814
DB 281 ARTAYACNGTNAARGTNGGNTNCARACNYTCCNGAYAAAYWSNACNWSNGARYNTNG 340
QY 815 CCAAGATCATCATATTGAATTCACCCCACTGTACCCCAAGACATGACATCGCCCTCA 874
DB 341 TNAONMGNATHGNTNATHCAYGARAAYTYYTHAAYMGNATGWSNGAYGAYATHGNAHY 400
QY 875 TGAAGCTGCACTTCCCACTCCTTTCTCAGGCACAGTACAGCCCATCTGTCTGCCCTTCT 934
DB 401 TNAARYTNAARTAYCCNGTNCACNTGWSNCCNYTNGTNCARCCNATHIGYTYNCCNWSNT 460
QY 935 TTGATGAGAGCTCACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAAC 994
DB 461 TYAAYTNAARCCNWSNATHGNGNACNATGTGTGGTNGTNGTNGGNGYNGARAARG 520
QY 995 AGAATGGA---GGGAAGATGTCTGACATCTGTGCGAGCGCTCAGTCCAGGTCAATTGACA 1051
DB 521 CNGARGNCAYCCNAARACNCCNTAYWSNGTNCARGNYTNGCNGTNGMGNATHGTAAYA 580
QY 1052 GCACCGTGTGCAATGCAGACGATGCT-----ACCAGGGGGAAGTCAACCG 1096
DB 581 AYGARATHGTAAAYCAYMGNATYCATTYTNTYNTYNTYNTYNTYNTYNTYNTYNTYNTY 640
QY 1097 AGAAGATGATGTGCGAGGCATCCCGGAAGGGGTGTGGACACCTGCCAGGTGACAGTG 1156
DB 641 GNAAYGAYATGYNTGYACNWSNWSNGARTGGGNYTNGAYACNTGYCARGAYACNWSNG 700
QY 1157 GTGGGCCCTGATGTACCAATCTGACACAG---TGGCATGTGTGGGCAATCGTTAGCTGG 1213
DB 701 GNWSNWSYNTNGTNGTCARATGAAYARACNTGGGTNCARATGGGNGTNGTNGTNGTNG 760
QY 1214 GCTATGCTGCGGGGCGCCAGCACCCAGGAGTATACACCAAGGTCTCAGGCTTATCTCA 1273
DB 761 AYTTCAYTGTGNGMGNMGNCACTTCCNWSNGTNTAYACNWSNACNWSNCAATYYACNC 820
QY 1274 ACTGGATCTA 1283
DB 821 ARTGGATHAA 830

Search completed: June 8, 2004, 02:22:27
Job time : 123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2004, 01:16:00 ; Search time 598 Seconds
(without alignments)
9955.526 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305

Sequence: 1 atggatctgacagtgatca.....atgctggaaggtgagctg 1305

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2995936 seqs, 2280998010 residues

Total number of hits satisfying chosen parameters: 5991872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA: *

```
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq2.*
14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303.4	99.9	2038	13	US-09-968-415-18
2	1303.4	99.9	2038	15	US-10-180-719-18
3	1303	99.8	1314	16	US-10-295-027-133
4	1303	99.8	1314	16	US-10-295-027-778
5	1303	99.8	1314	16	US-10-295-027-790
6	1303	99.8	1314	16	US-10-295-027-830
7	1303	99.8	1314	16	US-10-295-027-979
8	1303	99.8	1314	16	US-10-173-399-88
9	1303	99.8	2081	9	US-09-851-588-7
10	1303	99.8	2307	15	US-10-097-340-317
11	1303	99.8	2307	15	US-10-171-311-217
12	1298.2	99.5	2137	10	US-09-776-191-3
13	1298.2	99.5	2137	15	US-10-156-214A-3
14	1294	99.2	2165	15	US-10-101-510-634
					Sequence 18, Appl
					Sequence 18, Appl
					Sequence 133, App
					Sequence 778, App
					Sequence 790, App
					Sequence 830, App
					Sequence 979, App
					Sequence 88, Appl
					Sequence 7, Appl
					Sequence 317, App
					Sequence 217, App
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 634, App

```
15 1292 99.0 2079 9 US-09-851-588-5
16 1292 99.0 2079 10 US-09-776-191-71
17 1292 99.0 2079 15 US-10-264-820-22
18 1292 99.0 2079 15 US-10-254-289-1
19 1292 99.0 2079 16 US-10-156-214A-38
20 1263 96.8 2063 10 US-09-888-257A-2
21 1263 96.8 2063 10 US-09-946-374-274
22 1263 96.8 2063 12 US-10-015-395A-274
23 1263 96.8 2063 13 US-10-206-915-329
24 1263 96.8 2063 13 US-10-199-670-329
25 1263 96.8 2063 13 US-10-201-858-329
26 1263 96.8 2063 13 US-10-208-890-329
27 1263 96.8 2063 13 US-10-208-034-329
28 1263 96.8 2063 13 US-10-201-853-329
29 1263 96.8 2063 13 US-10-063-745-111
30 1263 96.8 2063 13 US-10-063-512-111
31 1263 96.8 2063 13 US-10-063-513-111
32 1263 96.8 2063 13 US-10-063-549-111
33 1263 96.8 2063 13 US-10-063-589-111
34 1263 96.8 2063 13 US-10-063-551-111
35 1263 96.8 2063 13 US-10-174-581-329
36 1263 96.8 2063 13 US-10-176-749-329
37 1263 96.8 2063 13 US-10-176-749-329
38 1263 96.8 2063 13 US-10-176-914-329
39 1263 96.8 2063 13 US-10-176-915-329
40 1263 96.8 2063 13 US-10-006-485A-274
41 1263 96.8 2063 13 US-10-013-907A-274
42 1263 96.8 2063 13 US-10-015-499A-274
43 1263 96.8 2063 13 US-10-063-555-111
44 1263 96.8 2063 13 US-10-063-563-111
45 1263 96.8 2063 13 US-10-063-594-111
```

ALIGNMENTS

```
RESULT 1
US-09-968-415-18
; Sequence 18, Application US/09968415
; Publication No. US20020086334A1
; GENERAL INFORMATION:
```

APPLICANT: Bandman, Olga

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSER: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/968,415

FILING DATE: 26-Sep-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/659,151

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheela

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNOT13

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-968-415-18

Query Match 99.9%; Score 1303.4; DB 13; Length 2038;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTTCGATGTCAAAACCCCTCGCGAAACCC 60
DB 200 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTTCGATGTCAAAACCCCTCGCGAAACCC 259

QY 61 CQTATCCCATGAGGACCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CGTATCCCATGAGGACCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 319

QY 121 CTGGCGAGTATCATATGTTGTTCTCTCATCAAGGTGATTCGGATAAATACTACTTC 180
DB 320 CTGGCGAGTATCATATGTTGTTCTCTCATCAAGGTGATTCGGATAAATACTACTTC 379

QY 181 CTCTCGGGGACCTCTCCATTCATCCCGAGGAGCAGCTGTGACGAGAGCTGGAC 240
DB 380 CTCTCGGGGACCTCTCCATTCATCCCGAGGAGCAGCTGTGACGAGAGCTGGAC 439

QY 241 TGTCTCTTGGGGAGGAGGAGGACACTGTGTCAAGAGCTTCCCGAGGAGGCTGCGAGTG 300
DB 440 TGTCTCTTGGGGAGGAGGAGGAGGACACTGTGTCAAGAGCTTCCCGAGGAGGCTGCGAGTG 499

QY 301 CGAGTCCGCTCTTCCAAAGGACCGATCCACACTCGAGGTGTGGACTCGGCCACACAGGGAAC 360
DB 500 CGAGTCCGCTCTTCCAAAGGACCGATCCACACTCGAGGTGTGGACTCGGCCACACAGGGAAC 559

QY 361 TGGTCTCTGCTGTTTTCGACAACTTCACAGAGCTCTCGTTCAGAGCTGTAGGACG 420
DB 560 TGGTCTCTGCTGTTTTCGACAACTTCACAGAGCTCTCGTTCAGAGCTGTAGGACG 619

QY 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTCGCCACAGACGAGATCTG 480
DB 620 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTCGCCACAGACGAGATCTG 679

QY 481 GATGTTGTTGAAATCACAGAAACAGCAGAGCTTCGCATCGCGAATCAAGTGGGCCC 540
DB 680 GATGTTGTTGAAATCACAGAAACAGCAGAGCTTCGCATCGCGAATCAAGTGGGCCC 739

QY 541 TGTCTCTCAGGCTCCCTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 740 TGTCTCTCAGGCTCCCTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 799

QY 601 CCCCCTGTGGTGGGAGAGGAGGCTCTGTGGATTCCTGGCCCTTGGCAGGCTGAGCATC 660
DB 800 CCCCCTGTGGTGGGAGAGGAGGCTCTGTGGATTCCTGGCCCTTGGCAGGCTGAGCATC 859

QY 661 CAGTACGACAAACAGCAGCTGTGGGAGGAGGAGCTTCCTGGACCCCTGGTGGTCTCAGC 720
DB 860 CAGTACGACAAACAGCAGCTGTGGGAGGAGGAGCTTCCTGGACCCCTGGTGGTCTCAGC 919

QY 721 CGAGGCCACTGTCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCA 780
DB 920 CGAGGCCACTGTCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCA 979

QY 781 GCAAACTGGGAGGCTTCCCATCCCTGCTGCTGGGCAAGATCATCATTAATTAATCAAC 840
DB 980 GCAAACTGGGAGGCTTCCCATCCCTGCTGCTGGGCAAGATCATCATTAATTAATCAAC 1039

QY 841 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTCGAGTCCCACTCTCTTC 900
DB 1040 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTCGAGTCCCACTCTCTTC 1099

QY 901 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 960
DB 1100 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1159

QY 961 CCACTCTGGATCATTTGGATGGGCTTTAGAAAGAGATGGAGGAGATGTCTGACATA 1020
DB 1160 CCACTCTGGATCATTTGGATGGGCTTTAGAAAGAGATGGAGGAGATGTCTGACATA 1219

QY 1021 CTCTCGAGGCTGAGTCCAGGTCATTGACAGCACACACGCTGCAATGCAGACGATGCTAC 1080
DB 1220 CTCTCGAGGCTGAGTCCAGGTCATTGACAGCACACACGCTGCAATGCAGACGATGCTAC 1279

QY 1081 CAGGGGAAAGTCAACGAGAAAGATGTGTGAGGATCCCGAAAGGGGTGTGACACC 1140
DB 1280 CAGGGGAAAGTCAACGAGAAAGATGTGTGAGGATCCCGAAAGGGGTGTGACACC 1339

QY 1141 TGCACGSGTGCAGTGGTGGGCCCTGTGATGTACCAATCTGACAGTGGCATGTGTGGGC 1200
DB 1340 TGCACGSGTGCAGTGGTGGGCCCTGTGATGTACCAATCTGACAGTGGCATGTGTGGGC 1399

QY 1201 ATCGTTAGCTGGGGCTATGGCTGCGGGGCCCGAGCACCCCGAGGATATACCAAGGTC 1260
DB 1400 ATCGTTAGCTGGGGCTATGGCTGCGGGGCCCGAGCACCCCGAGGATATACCAAGGTC 1459

QY 1261 TCAGCCTATCTCAACTGGATCTCAATGTCTGAGGCTGAGCTG 1305
DB 1460 TCAGCCTATCTCAACTGGATCTCAATGTCTGAGGCTGAGCTG 1504

RESULT 2

US-10-180-719-18

; Sequence 18, Application US/10180719

; Publication No. US20030166246A1

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; Hillman, Jennifer L.

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Purvi

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/180,719

; FILING DATE: 25-Jun-2002

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/008,271

; FILING DATE: 16-Jan-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-180-719-18

Query Match 99.9%; Score 1303.4; DB 15; Length 2038;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAAAACC 60
Db ATGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAAAACC 259

QY 61 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
Db CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 319

QY 121 CTGCGAGTATCATCATGTGTTCTCTCATCAAGTGTATCTGGATAAATACTACTTC 180
Db CTGCGAGTATCATCATGTGTTCTCTCATCAAGTGTATCTGGATAAATACTACTTC 379

QY 181 CTCTGGCGGACCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGCTGGAC 240
Db CTCTGGCGGACCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGCTGGAC 439

QY 241 TGTCCCTTGGGGAGGAGCAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTG 300
Db TGTCCCTTGGGGAGGAGCAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTG 499

QY 301 CGAGTCCGCTCTCGAAGGACGATCAGCTGAGCTGTGAGCTGTGAGCTGCGGACAGGAAC 360
Db CGAGTCCGCTCTCGAAGGACGATCAGCTGAGCTGTGAGCTGTGAGCTGCGGACAGGAAC 559

QY 361 TGGTCTCTGCTGCTTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCCTGTAGCAG 420
Db TGGTCTCTGCTGCTTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCCTGTAGCAG 619

QY 421 ATGGGCTACAGCAAAACCCATTTTCAAGCTGTGAGATGTGCCAGACCAAGATCTG 480
Db ATGGGCTACAGCAAAACCCATTTTCAAGCTGTGAGATGTGCCAGACCAAGATCTG 679

QY 481 GATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGATCGCGAATCTAAGTGGGCC 540
Db GATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGATCGCGAATCTAAGTGGGCC 739

QY 541 TGTCTCTCAGGCTCCCTGCTCTCCTGCACTGTCTTGGCTGTGGGAGAGCCTGAAGACC 600
Db TGTCTCTCAGGCTCCCTGCTCTCCTGCACTGTCTTGGCTGTGGGAGAGCCTGAAGACC 799

QY 601 CCCGTGTGTTGGTGGGAGAGGCTCTGTGATTTTGGCTTGGCTTGGCAGTCAAGATC 660
Db CCCGTGTGTTGGTGGGAGAGGCTCTGTGATTTTGGCTTGGCTTGGCAGTCAAGATC 859

QY 661 CAGTACGCAAAACAGCAGCTGTGGAGGAGCATCTTGGACCCCTCTGCTCTCAGC 720
Db CAGTACGCAAAACAGCAGCTGTGGAGGAGCATCTTGGACCCCTCTGCTCTCAGC 919

QY 721 GCAGCCCACTGCTTCAAGAAACATACCGATGTGTTCAACTGGAAGGTGGGGAGGTCA 780
Db GCAGCCCACTGCTTCAAGAAACATACCGATGTGTTCAACTGGAAGGTGGGGAGGTCA 979

QY 781 GACAACTGGGAGCTTCCCATCCTGCTGTGCGCAAGATCATCATCATTTGAATTCAC 840
Db GACAACTGGGAGCTTCCCATCCTGCTGTGCGCAAGATCATCATCATTTGAATTCAC 1039

QY 841 CCCATGTACCCCAAGAACATCATCGCCCTCATGAAGCTGCAGTTCCTCACTCACTTC 900

RESULT 3
US-10-295-027-133
Sequence 133, Application US/10295027
Publication No. US2003023350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM.

QY 1040 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTTC 1099
QY 901 TAGGCACAGTACAGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 960
Db TAGGCACAGTACAGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACC 1159

QY 961 CCACTCTGGATCATTTGGATGGGCTTTTACAAAGACAGAAATGGAGGGAAGATGTCTGACATA 1020
Db CCACTCTGGATCATTTGGATGGGCTTTTACAAAGACAGAAATGGAGGGAAGATGTCTGACATA 1219

QY 1021 CTGCTCAGGCGTCACTCCAGGTCAATTTGACACACACCGTGCATGCAATGACACGATGCGTAC 1080
Db CTGCTCAGGCGTCACTCCAGGTCAATTTGACACACACCGTGCATGCAATGACACGATGCGTAC 1279

QY 1081 CAGGGGGAAGTACCCAGAGAGATGATGTGAGGATCCCGAAGGGGGTGTGGACACC 1140
Db CAGGGGGAAGTACCCAGAGAGATGATGTGAGGATCCCGAAGGGGGTGTGGACACC 1339

QY 1141 TGCCAGGCTGACAGTGGTGGGCGCCCTGATGTACCAATCTGACACAGTGGCATGTGTGGGCG 1200
Db TGCCAGGCTGACAGTGGTGGGCGCCCTGATGTACCAATCTGACACAGTGGCATGTGTGGGCG 1399

QY 1201 ATGTTAGTGTGGGCTATGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGTTC 1260
Db ATGTTAGTGTGGGCTATGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGTTC 1459

QY 1261 TCAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
Db TCAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1504

TYPE: DNA
ORGANISM: Homo sapiens
US-10-295-027-778

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCACAAACCCCTGCGCAACCCCG 62
DB 9 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCACAAACCCCTGCGCAACCCCG 68

QY 63 TATCCCTGGAGAGCTTCAGAAAGGTGGGATCCCATCATAGCACTACTGAGCCT 122
DB 69 TATCCCTGGAGAGCTTCAGAAAGGTGGGATCCCATCATAGCACTACTGAGCCT 128

QY 123 GCGAGTATCATCAATTTGTTGCTCTCATCAAGGTGATCTTGGATAAATACTACTTCT 182
DB 129 GCGAGTATCATCAATTTGTTGCTCTCATCAAGGTGATCTTGGATAAATACTACTTCT 188

QY 183 CTGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGATG 242
DB 189 CTGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGATG 248

QY 243 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCTCGAGTGGC 302
DB 249 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCTCGAGTGGC 308

QY 303 AGTCCGCTCTCCAAAGGACCGATCCACACTGAGCTGTGGACTCGGCCACAGGAACTG 362
DB 309 AGTCCGCTCTCCAAAGGACCGATCCACACTGAGCTGTGGACTCGGCCACAGGAACTG 368

QY 363 GTTCTCTGCTGTTTGGCAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
DB 369 GTTCTCTGCTGTTTGGCAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 428

QY 423 GGGCTACAGCAACCACTTTTCAGAGCTGTGAGATTTGGCCAGACAGGATCTGGA 482
DB 429 GGGCTACAGCAACCACTTTTCAGAGCTGTGAGATTTGGCCAGACAGGATCTGGA 488

QY 483 TGTGTTGAATACAGAAACAGCAGGAGCTTGGATGGGAATCAAGTGGGCTG 542
DB 489 TGTGTTGAATACAGAAACAGCAGGAGCTTGGATGGGAATCAAGTGGGCTG 548

QY 543 TCTCTCAGCTCCCTGCTCTCCCTGCATCTCTGCTGCTGGAGAGCCTGAGACCCC 602
DB 549 TCTCTCAGCTCCCTGCTCTCCCTGCATCTCTGCTGCTGGAGAGCCTGAGACCCC 608

QY 603 CGGTGTGGTGGGAGGAGGCTCTGTGGATTTTGGCTTGGCAGGTGAGATCCA 662
DB 609 CGGTGTGGTGGGAGGAGGCTCTGTGGATTTTGGCTTGGCAGGTGAGATCCA 668

QY 663 GTACCAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCCCTCAGGC 722
DB 669 GTACCAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCCCTCAGGC 728

QY 723 AGCCCACTGCTTCAGAAACATACCAGTGTGTCAACTGGAAGTGGGGGAGGCTCAGA 782
DB 729 AGCCCACTGCTTCAGAAACATACCAGTGTGTCAACTGGAAGTGGGGGAGGCTCAGA 788

QY 783 CAACTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATTTGAATCAACCC 842
DB 789 CAACTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATTTGAATCAACCC 848

QY 843 CATGTACCCCAAGACAATGACATCGCCTCATGAAGCTGAGTTCACCTTCTTC 902
DB 849 CATGTACCCCAAGACAATGACATCGCCTCATGAAGCTGAGTTCACCTTCTTC 908

QY 903 AGGCACTGAGGCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCACCCC 962
DB 909 AGGCACTGAGGCCATCTGTCTGCCCTTCTTTGATGAGGAGTCACTCCAGCACCCC 968

QY 963 ACTCTGGATCATTTGGATGGGCTTTTACGAAGCAGATGGAGGAGATGTCTGACATACT 1022

RESULT 5

US-10-295-027-790
; Sequence 790, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 790
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-790

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTGACATGATCACTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCCCG 62
DB 9 GGATCTGACATGATCACTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCCCG 68
QY 63 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
DB 69 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123 GGGAGTATCATTTGGTGTCTCTCATCAAGTGTATCTGGAATAATTAATTAATTCCT 182
DB 129 GGGAGTATCATTTGGTGTCTCTCATCAAGTGTATCTGGAATAATTAATTAATTCCT 188
QY 183 CTGCGGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAATG 242
DB 189 CTGCGGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAATG 248
QY 243 TCCCTTGGGGAGGAGCAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 302
DB 249 TCCCTTGGGGAGGAGCAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 308
QY 303 AGTCGCGCTCTCCAGGACCGATCCACACTGAGCTGTGTGAGCTCGGACACAGGAACTG 362
DB 309 AGTCGCGCTCTCCAGGACCGATCCACACTGAGCTGTGTGAGCTCGGACACAGGAACTG 368
QY 363 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 422
DB 369 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 428
QY 423 GGGCTACAGAGAAACCACTTTGAGCTGTGTGAGATGTCGACAGCAGGATCTGGA 482
DB 429 GGGCTACAGAGAAACCACTTTGAGCTGTGTGAGATGTCGACAGCAGGATCTGGA 488
QY 483 TGTGTGTAATACAGAAACCACTTTGAGCTGTGTGAGATGTCGACAGCAGGATCTGGA 542
DB 489 TGTGTGTAATACAGAAACCACTTTGAGCTGTGTGAGATGTCGACAGCAGGATCTGGA 548
QY 543 TCTCTCAGGCTCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 549 TCTCTCAGGCTCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
QY 603 CCGTGTGCTGGTGGGAGGAGGCTCTGTGATTTCTGGCTTGGCAGGTCAGATCCA 662
DB 609 CCGTGTGCTGGTGGGAGGAGGCTCTGTGATTTCTGGCTTGGCAGGTCAGATCCA 668
QY 663 GTACGAAACACAGCCTCTGTGGAGGAGCAGCTGTGAGACCCCTCTGAGCTCAAGCC 722
DB 669 GTACGAAACACAGCCTCTGTGGAGGAGCAGCTGTGAGACCCCTCTGAGCTCAAGCC 728
QY 723 AGCCCACTGCTCAGAAACATACCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 782
DB 729 AGCCCACTGCTCAGAAACATACCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 788
QY 783 CAAACTGGGAGCTTCCCATCCCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
DB 789 CAAACTGGGAGCTTCCCATCCCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 848
QY 843 CATGTACCCCAAGAACATGATCCCTCTGAGCTGTGAGTCCCACTCACTTCTC 902
DB 849 CATGTACCCCAAGAACATGATCCCTCTGAGCTGTGAGTCCCACTCACTTCTC 908
QY 903 AGGCACAGTCAGGCCCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 962
DB 909 AGGCACAGTCAGGCCCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968
QY 963 ACTCTGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGAGAGATGCTGATCAT 1022
DB 969 ACTCTGATCATTTGGATGGGCTTTTACGAAGCAGATGAGGAGAGATGCTGATCAT 1028
QY 1023 GCTGACGGCGTCAGTCCAGTCTATTACAGACACACCGGTGCAATGCAGAGATGCGTACCA 1082

Db 1029 GCTGCAAGGGCTCAGTCCAGTCAATGCAGACACACCGGTGCAATGCAGAGATGCGTACCA 1088
QY 1083 GGGGGAAGTCAACGAGAAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGACACCTG 1142
DB 1089 GGGGGAAGTCAACGAGAAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGACACCTG 1148
QY 1143 CAGGGTGAAGTGTGGGCTCTGATGTACCAATCTGACCACTGACATGTGCTGGGCT 1202
DB 1149 CAGGGTGAAGTGTGGGCTCTGATGTACCAATCTGACCACTGACATGTGCTGGGCT 1208
QY 1203 CATTAGTGGGCTGATGGCTGCGGGGCGCCAGCACCCAGGAGTATACCAAGGTCTC 1262
DB 1209 CATTAGTGGGCTGATGGCTGCGGGGCGCCAGCACCCAGGAGTATACCAAGGTCTC 1268
QY 1263 AGCTATCTCAATGATCTACCAATGCTTGGAAAGGCTGAGCTG 1305
DB 1269 AGCTATCTCAATGATCTACCAATGCTTGGAAAGGCTGAGCTG 1311

RESULT 6
US-10-295-027-830
; Sequence 830, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 830
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-830

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCCTGACAGTGTACCTCTGAACAGCCTCGATGTCAACCCCTCGCGAAACCCCG 62
Db 9 GGATCCTGACAGTGTACCTCTGAACAGCCTCGATGTCAACCCCTCGCGAAACCCCG 68
QY 63 TATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 69 TATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123 GCGAGATATCATATGTGGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 182
Db 129 GCGAGATATCATATGTGGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 188
QY 183 CTGCGGCGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGCACTG 242
Db 189 CTGCGGCGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGCACTG 248
QY 243 TCCCTTGGGGAGGACGAGGACATGTGTCAAGAGCTTCCCGAAGGCGCTGCACTGCG 302
Db 249 TCCCTTGGGGAGGACGAGGACATGTGTCAAGAGCTTCCCGAAGGCGCTGCACTGCG 308
QY 303 AGTCGCGCTCTCCAGGACCGATCCACTGCTGAGCTGCTGAGCTCGGCCACAGGAACTG 362
Db 309 AGTCGCGCTCTCCAGGACCGATCCACTGCTGAGCTGCTGAGCTCGGCCACAGGAACTG 368
QY 363 GTTCTGTGCTGTTTCGAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 422
Db 369 GTTCTGTGCTGTTTCGAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 428
QY 423 GGGCTACAGCAACCCACTTTCAGAGCTGTGAGATGCGCCACAGACAGGATCTGGA 482
Db 429 GGGCTACAGCAACCCACTTTCAGAGCTGTGAGATGCGCCACAGACAGGATCTGGA 488
QY 483 TGTGTGTGAATTCACAGAAACAGCAGGAGCTTCGCTGCGGAATCTCAAGTGGGCCCTG 542
Db 489 TGTGTGTGAATTCACAGAAACAGCAGGAGCTTCGCTGCGGAATCTCAAGTGGGCCCTG 548
QY 543 TCTCTCAGCTCCCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
Db 549 TCTCTCAGCTCCCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
QY 603 CCGTGTGTGGTGGGAGGAGCGCTCTGTGATTTCTGGCTTGGAGTGGAGCTCAGATCCA 662
Db 609 CCGTGTGTGGTGGGAGGAGCGCTCTGTGATTTCTGGCTTGGAGTGGAGCTCAGATCCA 668
QY 663 GTACGACAAACAGCAGCTCTGGAGGAGGAGCTCTGAGTCTTGGCTTGGAGTGGAGCTCAG 722
Db 669 GTACGACAAACAGCAGCTCTGGAGGAGGAGCTCTGAGTCTTGGCTTGGAGTGGAGCTCAG 728
QY 723 AGCCCACTGCTTCAGAAACATACGATGTGTCACTGGRAGTGGCGGAGGCTCAGA 782
Db 729 AGCCCACTGCTTCAGAAACATACGATGTGTCACTGGRAGTGGCGGAGGCTCAGA 788
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 842
Db 789 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 848
QY 843 CATGTACCCCAAGACATGATGCCCTCATGAAGCTGAGTTCCTCCACTCATCTTCTC 902
Db 849 CATGTACCCCAAGACATGATGCCCTCATGAAGCTGAGTTCCTCCACTCATCTTCTC 908
QY 903 AGGCAAGTCAGGCGCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCC 962
Db 909 AGGCAAGTCAGGCGCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCC 968
QY 963 ACTCTGGATCATGTAGTGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1022
Db 969 ACTCTGGATCATGTAGTGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1028
QY 1023 GCTGAGGCGTCACTCCAGGTCATTGACAGCACAGGTCGATGACAGAGTGGGTACCA 1082
Db 1029 GCTGAGGCGTCACTCCAGGTCATTGACAGCACAGGTCGATGACAGAGTGGGTACCA 1088
QY 1083 GGGGGAAGTCACCGAGAGATGATGTGTGAGGATCCCGGAAGGGGGTGTGGACACCTG 1142

Db 1089 GGGGGAAGTCACCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACTG 1148
QY 1143 CCAGGCTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAAGTGGCATGTGTGGGCAT 1202
Db 1149 CCAGGCTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCAAGTGGCATGTGTGGGCAT 1208
QY 1203 CGTTAGCTGGGCTATGGCTGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1262
Db 1209 CGTTAGCTGGGCTATGGCTGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC 1268
QY 1263 AGCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
Db 1269 AGCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1311

RESULT 7

US-10-295-027-979
; Sequence 979, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2000-09-15
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-11-15
; PRIOR FILING DATE: 2001-11-21
; PRIOR FILING DATE: 2001-11-29
; PRIOR FILING DATE: 2001-12-14
; PRIOR FILING DATE: 2002-01-08
; PRIOR FILING DATE: 2002-01-10
; PRIOR FILING DATE: 2002-02-08
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 979
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-979

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATCCTGACAGTGTACCTCTGAACAGCCTCGATGTCAACCCCTCGCGAAACCCCG 62
Db 9 GGATCCTGACAGTGTACCTCTGAACAGCCTCGATGTCAACCCCTCGCGAAACCCCG 68

QY 63 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 69 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123 GCGAGATATCATATGTGGTGTCTCATCAAGTGTATCTGGATATAAATACTACTTCT 182
Db 129 GCGAGATATCATATGTGGTGTCTCATCAAGTGTATCTGGATATAAATACTACTTCT 188
QY 183 CTGCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 242
Db 189 CTGCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 248
QY 243 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 302
Db 249 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 308
QY 303 AGTCGGCTCTCCAGGACCGATCCACACTGAGCTGTGGTGTGGTGTGGTGTGGTGTGG 362
Db 309 AGTCGGCTCTCCAGGACCGATCCACACTGAGCTGTGGTGTGGTGTGGTGTGGTGTGG 368
QY 363 GTTCTCTGCTGTCTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAGAT 422
Db 369 GTTCTCTGCTGTCTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAGAT 428
QY 423 GGGCTACAGAGCAAAACCCACTTTCAGAGCTGTGGAGATGGCCACAGCAGGATCTGGA 482
Db 429 GGGCTACAGAGCAAAACCCACTTTCAGAGCTGTGGAGATGGCCACAGCAGGATCTGGA 488
QY 483 TGTGTGTAATACAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCGCTG 542
Db 489 TGTGTGTAATACAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCGCTG 548
QY 543 TCTCTCAGGCTCCTGCTTCCCTGCACTGTCTGCTGTGGGAGAGCCTTGAAGACCCC 602
Db 549 TCTCTCAGGCTCCTGCTTCCCTGCACTGTCTGCTGTGGGAGAGCCTTGAAGACCCC 608
QY 603 CGGTGTGGTGGGAGGAGCCTCTGTGATTTCTGGCTTGGCCTTGGCAGTCAAGTCCA 662
Db 609 CGGTGTGGTGGGAGGAGCCTCTGTGATTTCTGGCTTGGCCTTGGCAGTCAAGTCCA 668
QY 663 GTACGACAAACAGCAGCTGTGGAGGAGCAGTCTGGAACCCCTGCTGCTTCAAGCC 722
Db 669 GTACGACAAACAGCAGCTGTGGAGGAGCAGTCTGGAACCCCTGCTGCTTCAAGCC 728
QY 723 AGCCCACTGTTTCAAGAAATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 782
Db 729 AGCCCACTGTTTCAAGAAATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 788
QY 783 CAAACTGGGAGCCTTCCCATCCCTGCTGTGGCCCAAGATCATCATTTGAATTCAACCC 842
Db 789 CAAACTGGGAGCCTTCCCATCCCTGCTGTGGCCCAAGATCATCATTTGAATTCAACCC 848
QY 843 CATGTACCCCAAGAAATACATCCCTCTCATGAGCTGAGTTCCTTCCACTTCACTTCTC 902
Db 849 CATGTACCCCAAGAAATACATCCCTCTCATGAGCTGAGTTCCTTCCACTTCACTTCTC 908
QY 903 AGGCACAGTCAGGCCCATCTGTCTGCTTCTTTGATGAGAGCTCACTTCAAGCACCCC 962
Db 909 AGGCACAGTCAGGCCCATCTGTCTGCTTCTTTGATGAGAGCTCACTTCAAGCACCCC 968
QY 963 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGATGAGGAGAGATGTCTGATATCT 1022
Db 969 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGATGAGGAGAGATGTCTGATATCT 1028
QY 1023 GCTGAGGGGTGAGTCCAGTCAATTCACAGACACCGGTGCAATGAGAGAGTGGCTACCA 1082
Db 1029 GCTGAGGGGTGAGTCCAGTCAATTCACAGACACCGGTGCAATGAGAGAGTGGCTACCA 1088
QY 1083 GGGGGAAGTACCGAGAAATGATGTGTGAGGATCCCGAAGGGGGTGTGACACCTG 1142
Db 1089 GGGGGAAGTACCGAGAAATGATGTGTGAGGATCCCGAAGGGGGTGTGACACCTG 1148
QY 1143 CAGGGGTGACATGGTGGGGCCCTGTATGTACCAATCTGACAGTGGCATGTGGTGGCAT 1202

Db 1149 CCAGGGTGACAGTGGTGGGCCCTGTATGTACCAATGTGACCAATGTGCTGGGCAT 1208
QY 1203 CGTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTTC 1262
Db 1209 CGTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTTC 1268
QY 1263 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGTGAGCTG 1305
Db 1269 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGTGAGCTG 1311

RESULT 8
US-10-173-999-88
; Sequence 88, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/372,246
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-173-999-88

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTCGCAAAACCCCG 62
Db 9 GGATCTTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTCGCAAAACCCCG 68
QY 63 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 69 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123 GCGAGTATCATATGTGGTGTCTCTCATCAAGTGTATCTGGATATAAATACTACTTCT 182
Db 129 GCGAGTATCATATGTGGTGTCTCTCATCAAGTGTATCTGGATATAAATACTACTTCT 188
QY 183 CTGCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 242
Db 189 CTGCGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 248
QY 243 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 302
Db 249 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 308
QY 303 AGTCGGCTCTCCAGGACCGATCCACACTGAGCTGTGGTGTGGTGTGGTGTGGTGTGG 362
Db 309 AGTCGGCTCTCCAGGACCGATCCACACTGAGCTGTGGTGTGGTGTGGTGTGGTGTGG 368
QY 363 GTTCTCTGCTGTCTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAGAT 422
Db 369 GTTCTCTGCTGTCTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAGAT 428

QY	423	GGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGGAGATTGGCCAGACCAAGATCTCGA	482
Db	429	GGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGGAGATTGGCCAGACCAAGATCTCGA	488
QY	483	TGTTGTTGAAATCAGAAAAACAGCAGGAGCTTCGCATGCGGAACTCAAGTGGGCCCTG	542
Db	489	TGTTGTTGAAATCAGAAAAACAGCAGGAGCTTCGCATGCGGAACTCAAGTGGGCCCTG	548
QY	543	TCTCTCAGGCTCCCTGGTCTCCCTGCACATGCTTCCTGTGGAGAGAGCCTGMAAGACCC	602
Db	549	TCTCTCAGGCTCCCTGGTCTCCCTGCACATGCTTCCTGTGGAGAGAGCCTGMAAGACCC	608
QY	603	CCGTGTGGTGGGTGGGAGAGGCCTCTGTGGAATCTTTGGCTTGGCAGGTGAGCATCCA	662
Db	609	CCGTGTGGTGGGTGGGAGAGGCCTCTGTGGAATCTTTGGCTTGGCAGGTGAGCATCCA	668
QY	663	GTACGACAAACAGCAGCTCTGTGGAGGAGCATCCTGCAACCCCACTGGGTCTCTCAGGC	722
Db	669	GTACGACAAACAGCAGCTCTGTGGAGGAGCATCCTGCAACCCCACTGGGTCTCTCAGGC	728
QY	723	AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGA	782
Db	729	AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGA	788
QY	783	CAAACTGGGCACTTCCATCCCTGGCTGTGGCCAGATCATCATATGTAATCAACCC	842
Db	789	CAAACTGGGCACTTCCATCCCTGGCTGTGGCCAGATCATCATATGTAATCAACCC	848
QY	843	CATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTTCTC	902
Db	849	CATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTTCTC	908
QY	903	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCATCCAGCCACCC	962
Db	909	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCATCCAGCCACCC	968
QY	963	ACTCTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACT	1022
Db	969	ACTCTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACT	1028
QY	1023	GCTGCAGGCTCAGTCCAGTTCATGTACAGCACACGGTGCAAATGCAGACGATGGGTACCA	1082
Db	1029	GCTGCAGGCTCAGTCCAGTTCATGTACAGCACACGGTGCAAATGCAGACGATGGGTACCA	1088
QY	1083	GGGGGAAGTCACCGAGAAGATCATGTGTGCAGGCAATCCGGAAGGGGTGTGGACACCTG	1142
Db	1089	GGGGGAAGTCACCGAGAAGATCATGTGTGCAGGCAATCCGGAAGGGGTGTGGACACCTG	1148
QY	1143	CCAGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACAGTGGCAATGTGGTGGGCAT	1202
Db	1149	CCAGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACAGTGGCAATGTGGTGGGCAT	1208
QY	1203	CGTTAGCTGGGCTATGCTCGGGGGCCGAGCACCCAGGAGTATACCAAGAGTCTC	1262
Db	1209	CGTTAGCTGGGCTATGCTCGGGGGCCGAGCACCCAGGAGTATACCAAGAGTCTC	1268
QY	1263	AGCCTATCTCAACTGGATCTCAATGTCTGAAAGGCTGAGCTG	1305
Db	1269	AGCCTATCTCAACTGGATCTCAATGTCTGAAAGGCTGAGCTG	1311

RESULT 9

US-09-851-588-7
; Sequence 7, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JUD/AMS

QY	783	CAAACTGGGCAGCTTCCCATCCCTGGCTGGGCCAAGATCATCATGTAATTCAAACC	842
Db	1003	CAAACTGGGCAGCTTCCCATCCCTGGCTGGGCCAAGATCATCATGTAATTCAAACC	1062
QY	843	CATGTATCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTC	902
Db	1063	CATGTATCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTCTC	1122
QY	903	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCC	962
Db	1123	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCC	1182
QY	963	ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGAAGATGTTGCATACT	1022
Db	1183	ACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGAAGATGTTGCATACT	1242
QY	1023	GCTGCAGGCGTCACTCCAGTCAATTGACAGCACACGGTGCAATGCAGACGATCGGTACCA	1082
Db	1243	GCTGCAGGCGTCACTCCAGTCAATTGACAGCACACGGTGCAATGCAGACGATCGGTACCA	1302
QY	1083	GGGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGCACACCTG	1142
Db	1303	GGGGGAAGTCACCGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGCACACCTG	1362
QY	1143	CCAGGGTGAACGTGGTGGGCCCTTGATGTATCAATCTGACCACTGGCATGTGTGGGCAT	1202
Db	1363	CCAGGGTGAACGTGGTGGGCCCTTGATGTATCAATCTGACCACTGGCATGTGTGGGCAT	1422
QY	1203	CGTTAGCTGGGGCTATGGCTGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC	1262
Db	1423	CGTTAGCTGGGGCTATGGCTGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTC	1482
QY	1263	AGCCTATCTCAACTGGATCTACAATGTTCTGGAAGGCTGAGCTG	1305
Db	1483	AGCCTATCTCAACTGGATCTACAATGTTCTGGAAGGCTGAGCTG	1525

RESULT 10

RES001 10
US-10-097-340-317
; Sequence 317, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAWATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHWANDT
; APPLICANT: Xumel ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732

QY 783 CAAATGGGCGAGTCCCATCCCTGGCTGTGGCCAGATCATCATTAATTAACCC 842
Db 1063 CAAATGGGCGAGTCCCATCCCTGGCTGTGGCCAGATCATCATTAATTAACCC 1122
QY 843 CATGTACCCCAAGACAATAGACATCGCCCTCATGAGCTGCAATTTCCCACTACATTTCTC 902
Db 1123 CATGTACCCCAAGACAATAGACATCGCCCTCATGAGCTGCAATTTCCCACTACATTTCTC 1182
QY 903 AGGCAGTACGCGCCCATCTGTCTGCCCTTTTACGAAGCAGAAATGAGGAAAGATGTCTGCATACT 1022
Db 1183 AGGCAGTACGCGCCCATCTGTCTGCCCTTTTACGAAGCAGAAATGAGGAAAGATGTCTGCATACT 1302
QY 963 ACTCTGGATCATTTGATGGGGCTTTTACGAAGCAGAAATGAGGAAAGATGTCTGCATACT 1082
Db 1243 ACTCTGGATCATTTGATGGGGCTTTTACGAAGCAGAAATGAGGAAAGATGTCTGCATACT 1362
QY 1023 GCTGAGGGCTGAGTCCAGGTCAATGACAGACACAGGTGCAATGACAGAGTGGGTACCA 1082
Db 1303 GCTGAGGGCTGAGTCCAGGTCAATGACAGACACAGGTGCAATGACAGAGTGGGTACCA 1362
QY 1083 GGGGAAGTCAACGAGAGATGATGTGAGGCAATCCCGAAGGGGTGTGGACACTG 1142
Db 1363 GGGGAAGTCAACGAGAGATGATGTGAGGCAATCCCGAAGGGGTGTGGACACTG 1422
QY 1143 CCAGGTGACAGTGTGGGCCCCCTCATGTACCAATCTGACAGTGGCATGTGGGCA 1202
Db 1423 CCAGGTGACAGTGTGGGCCCCCTCATGTACCAATCTGACAGTGGCATGTGGGCA 1482
QY 1203 CTTAGCTGGGCTATGCTGTGGGGCCCCGAGCACCCAGGATATACCAAGGTTCT 1282
Db 1483 CTTAGCTGGGCTATGCTGTGGGGCCCCGAGCACCCAGGATATACCAAGGTTCT 1542
QY 1263 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
Db 1543 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1585

RESULT 11

US-10-171-311-217
; Sequence 217, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerssh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217
; LENGTH: 2307
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 1693, 1598, 1705, 1708, 1709, 1713, 1717, 1720, 1724, 1728,
; LOCATION: 1733, 1741, 1746, 1748, 1755, 1770, 1774, 1791, 1802, 1821,
; LOCATION: 1838, 1856, 1859, 1864, 1908, 1959, 2012, 2038, 2143

; OTHER INFORMATION: n = A, T, C or G
US-10-171-311-217

Query Match 99.8%; Score 1303; DB 15; Length 2307;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 62
Db 283 GGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 342
QY 63 TATCCCATCGAGACCTTTCAAAAGGTGGGATCCCCATCATATAGCACTACTAGACCT 122
Db 343 TATCCCATCGAGACCTTTCAAAAGGTGGGATCCCCATCATATAGCACTACTAGACCT 402
QY 123 GGGAGTATCATATTTGTTGCTTCTCAAGGTGATTTCTGGATAAATPACTTTCT 182
Db 403 GGGAGTATCATATTTGTTGCTTCTCAAGGTGATTTCTGGATAAATPACTTTCT 462
QY 183 CTGGGGCAGCTCTCCACTTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGTGGACTG 242
Db 463 CTGGGGCAGCTCTCCACTTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGTGGACTG 522
QY 243 TCCTTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCGAGTGGC 302
Db 523 TCCTTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCGAGTGGC 582
QY 303 AGTCGGCTCTCCAAGACCGATCCACACTGCAAGTGTCTGGACTCGGCCACAGGAACTG 362
Db 583 AGTCGGCTCTCCAAGACCGATCCACACTGCAAGTGTCTGGACTCGGCCACAGGAACTG 642
QY 363 GTTCTCTGCTGTTTTCGACAACTTTCAGAAAGCTCTCGCTGAGACAGCCTGTAGGAGAT 422
Db 643 GTTCTCTGCTGTTTTCGACAACTTTCAGAAAGCTCTCGCTGAGACAGCCTGTAGGAGAT 702
QY 423 GGGCTACAGCAGCAAAACCCACTTTTCAAGCTGTGGAGATTTGGCCACAGATCTGGA 482
Db 703 GGGCTACAGCAGCAAAACCCACTTTTCAAGCTGTGGAGATTTGGCCACAGATCTGGA 762
QY 483 TGTGTTGAAATCACAGAAACAGCCAGGAGCTTCCGATCGGAACTCAAGTGGGCCCCG 542
Db 763 TGTGTTGAAATCACAGAAACAGCCAGGAGCTTCCGATCGGAACTCAAGTGGGCCCCG 822
QY 543 TCTCTAGGCTCCCTGGTCTCCTGCACTGTCTTGTGCTGTGGAAAGAGCTGAAGACCCC 602
Db 823 TCTCTAGGCTCCCTGGTCTCCTGCACTGTCTTGTGCTGTGGAAAGAGCTGAAGACCCC 882
QY 603 CCGTGTGGTGGGAGGAGGAGCCCTCTGTGGATTTCTTGGCTTGGCAGTCAAGATCCA 662
Db 883 CCGTGTGGTGGGAGGAGGAGCCCTCTGTGGATTTCTTGGCTTGGCAGTCAAGATCCA 942
QY 663 GTACGACAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCACTGGGTCTCAGGC 722
Db 943 GTACGACAAACAGCAGCTCTGTGGAGGAGCAGCTCTGGACCCCACTGGGTCTCAGGC 1002
QY 723 AGCCCACTGTTTCAAGAAACATACCGATGTGTTCATCTGGAAGGTGCGGAGGCTCAGA 782
Db 1003 AGCCCACTGTTTCAAGAAACATACCGATGTGTTCATCTGGAAGGTGCGGAGGCTCAGA 1062
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCC 842
Db 1063 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCC 1122
QY 843 CATGTACCCCAAGACAATAGATCCCTCATGAAGCTGCAATTTCCCACTCACTTTCTC 902
Db 1123 CATGTACCCCAAGACAATAGATCCCTCATGAAGCTGCAATTTCCCACTCACTTTCTC 1182
QY 903 AGGCAGTACGCGCCCATCTGTCTGCCCTTTTACGAAGGAGGCTCACTCCAGCCACCCC 962
Db 1183 AGGCAGTACGCGCCCATCTGTCTGCCCTTTTACGAAGGAGGCTCACTCCAGCCACCCC 1242
QY 963 ACTCTGATCATTTGATGGGCTTTTACGAAGCAGAAATGAGGAAAGATGTCTGACATCT 1022

1243 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAATGGAGGAAGATGTTCTGACATACT 1302
1023 GGTGAGGGGTGAGTCCAGTCAATTCAGACACAGCGTGGCAATGCAGACGATGGGTACCA 1082
1303 GGTGAGGGGTGAGTCCAGTCAATTCAGACACAGCGTGGCAATGCAGACGATGGGTACCA 1362
1083 GGGGGAAGTCAACGAGAGATGATGTCAGGAGATCCCGGAAGGGGGTGTGGACACCTG 1142
1363 GGGGGAAGTCAACGAGAGATGATGTCAGGAGATCCCGGAAGGGGGTGTGGACACCTG 1422
1143 CCAGGTGACATGTTGGGCCCCCTGATGATCAATCTGACAGTGGCATGTTGGGGCAT 1202
1423 CCAGGTGACATGTTGGGCCCCCTGATGATCAATCTGACAGTGGCATGTTGGGGCAT 1482
1203 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATACACCAAGGTCTC 1262
1483 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATACACCAAGGTCTC 1542
1263 AGCTATCTCAACTGGATCTACAATGTTCTGGAAGGCTGAGCTG 1305
1543 AGCTATCTCAACTGGATCTACAATGTTCTGGAAGGCTGAGCTG 1585

RESULT 12
US-09-776-191-3
; Sequence 3, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Chn
; APPLICANT: Jiumn-Chern Yeh
; TITLE OF INVENTION: CORVAS INTERNATIONAL, INC.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2137
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)...(1574)
; OTHER INFORMATION: DNA sequence encoding a transmembrane serine
; OTHER INFORMATION: protease (MTSP3) protein

US-09-776-191-3
Query Match 99.5%; Score 1298.2; DB 10; Length 2137;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
3 GGATCTGACAGTGAATCACTCTGAACAGCCTGATGTCAAAACCCCTGCGAAACCCCG 62
269 GGATCTGACAGTGAATCACTCTGAACAGCCTGATGTCAAAACCCCTGCGAAACCCCG 328
63 TATCCCCATGGACACCTTCAGAAAGTGGGGATCCCATCATCAGTACTGAGGCT 122
329 TATCCCCATGGACACCTTCAGAAAGTGGGGATCCCATCATCAGTACTGAGGCT 388

123 GGCACTATCATCATTTGGTGTCTCTCATCAAGGTGATTTCTGGATTAATACTACTTCTT 182
389 GCGAGTATCATCATTTGGTGTCTCTCATCAAGGTGATTTCTGGATTAATACTACTTCTT 448
183 CTGGGGGAGCCTCTCACTTCCATCCGAGGAGAGAGCTGTGTGACGAGAGCTGAGCTG 242
449 CTGGGGGAGCCTCTCTCATTTCCATCCGAGGAGAGAGCTGTGTGACGAGAGCTGAGCTG 508
243 TCCCTTTGGGGAGGAGAGAGAGCACTGTGTCAAGAGCTTTCCCGAAGGCGCTGCAAGTGGC 302
509 TCCCTTTGGGGAGGAGAGAGAGCACTGTGTCAAGAGCTTTCCCGAAGGCGCTGCAAGTGGC 568
303 AGTCCCGCTCTCAAGAGAGAGAGTCCACACTGAGGAGCTGTGTGAGCTCGGCCACAGGAGCTG 362
569 AGTCCCGCTCTCAAGAGAGAGAGTCCACACTGAGGAGCTGTGTGAGCTCGGCCACAGGAGCTG 628
363 GTTCTCTGCGCTGTTTCGACAACTTTCAGAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
629 GTTCTCTGCGCTGTTTCGACAACTTTCAGAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 688
423 GGGCTACAGCAGCAAAACCCCACTTTCAGAGAGCTGTGTGAGAGTGTGAGAGAGCTGTGAG 482
689 GGGCTACAGCAGCAAAACCCCACTTTCAGAGAGCTGTGTGAGAGTGTGAGAGAGCTGTGAG 748
483 TGTGTTGTAATCACAGAAACAGCCAGGAGCTTCCATGCGGAATCAAGTGGGCGCTG 542
749 TGTGTTGTAATCACAGAAACAGCCAGGAGCTTCCATGCGGAATCAAGTGGGCGCTG 808
543 TCTCTCAGGCTCCCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
809 TCTCTCAGGCTCCCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 868
603 CCGTGTGGTGGTGGGAGAGAGCTCTGTGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 662
869 CCGTGTGGTGGTGGGAGAGAGCTCTGTGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 928
663 GTACGACAAACAGCAGCTCTGTGAGGAGAGCTCTGTGAGGAGAGCTCTGTGAGGAGAGCTCT 722
929 GTACGACATACAGCAGCTCTGTGAGGAGAGCTCTGTGAGGAGAGCTCTGTGAGGAGAGCTCT 988
723 AGCCCACTGCTCAGGAAACATACCATGTTTCAACTGGAAGTGTGGGAGAGCTCTGAG 782
989 AGCCCACTGCTCAGGAAACATACCATGTTTCAACTGGAAGTGTGGGAGAGCTCTGAG 1048
783 CAACACTGGCAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
1049 CAACACTGGCAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1108
843 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTGTCCACTTCTTCTC 902
1109 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTGTCCACTTCTTCTC 1168
903 AGCAGAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962
1169 AGCAGAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
963 ACTCTGATCATTTGGATGGGCTTTACGAAGAGAGATGGAGGAGAGATGTCTGACATCT 1022
1229 ACTCTGATCATTTGGATGGGCTTTACGAAGAGAGATGGAGGAGAGATGTCTGACATCT 1288
1023 GCTCAGGCTGAGTCCAGGTCTTGTGACAGCAGCAGCTGCAATGACAGAGATGCTGACATCT 1082
1289 GCTCAGGCTGAGTCCAGGTCTTGTGACAGCAGCAGCTGCAATGACAGAGATGCTGACATCT 1348
1083 GGGGGAAGTCAACGAGAGAGATGATGTGAGGAGATCCCGGAAGGGGGTGTGACACCTG 1142
1349 GGGGGAAGTCAACGAGAGAGATGATGTGAGGAGATCCCGGAAGGGGGTGTGACACCTG 1408
1143 CCAGGTGACAGTGGTGGGCCCCCTGATGATCAATCTGACAGTGGCATGTTGGGGCAT 1202
1409 CCAGGTGACAGTGGTGGGCCCCCTGATGATCAATCTGACAGTGGCATGTTGGGGCAT 1468
1203 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATACACCAAGGTCTC 1562

Db 1469 CGTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCCAGGATATACACCAAGGTCTC 1528
Qy 1263 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1305
Db 1529 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1571

RESULT 13
US-10-156-214A-3
; Sequence 3, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Siev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
; FILE OF INVENTION: Theresof
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2137
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)...(1574)
; OTHER INFORMATION: Nucleic acid encoding a transmembrane serine
; OTHER INFORMATION: protease (MTP23) protein
US-10-156-214A-3

Query Match 99.5%; Score 1298.2; DB 16; Length 2137;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 62
Db 269 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 328
Qy 53 TATCCCGATGAGACCTTCAGAAAGTGGGATCCCATCATAGCACTACTAGGCT 122
Db 323 TATCCCGATGAGACCTTCAGAAAGTGGGATCCCATCATAGCACTACTAGGCT 388
Qy 123 GCGAGTATCATATGTGGTGTCTCATCAAGTGTATCTGGATAAATACTTCTCT 182
Db 389 GCGAGTATCATATGTGGTGTCTCATCAAGTGTATCTGGATAAATACTTCTCT 448
Qy 183 CTGCGGGAGGCTCTTCCATCTTCATCCGAGGAGCAGCTGTGTGACCGGAGAGCTGGA 242
Db 449 CTGCGGGAGGCTCTTCCATCTTCATCCGAGGAGCAGCTGTGTGACCGGAGAGCTGGA 508
Qy 243 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGGCTTCCCGAAGGCTGCGAGTGGC 302
Db 509 TCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGGCTTCCCGAAGGCTGCGAGTGGC 568
Qy 303 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCTGAGCTGCGGACACAGGGAATG 362
Db 569 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCTGAGCTGCGGACACAGGGAATG 628
Qy 363 GTTCTCTGCTGTTTCGCAACTTCAGAAAGCTTCGCTGAGACAGCTTAGGCAGAT 422
Db 629 GTTCTCTGCTGTTTCGCAACTTCAGAAAGCTTCGCTGAGACAGCTTAGGCAGAT 688
Qy 423 GGGCTTACAGCAAAACCCACTTTTCAGAGCTGTGAGATTGGCCACAGCAAGGATCTGGA 482
Db 689 GGGCTTACAGCAAAACCCACTTTTCAGAGCTGTGAGATTGGCCACAGCAAGGATCTGGA 748

RESULT 14

US-10-101-510-634
; Sequence 634, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 634

Qy 483 TGTGTTGAAATCACAGAAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCGCTG 542
Db 749 TGTGTTGAAATCACAGAAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCGCTG 808
Qy 543 TCTCTCAGGCTCCCTGCTCTCCCTGCACTCTCTGCTGCTGTGGGAGAGCTGAGACCCC 602
Db 809 TCTCTCAGGCTCCCTGCTCTCCCTGCACTCTCTGCTGCTGTGGGAGAGCTGAGACCCC 868
Qy 603 CCGTGTGCTGGTGGGAGGAGGCTCTGCTGGGATCTTGGCCCTGGCAGGTGAGCATCCA 662
Db 869 CCGTGTGCTGGTGGGAGGAGGCTCTGCTGGGATCTTGGCCCTGGCAGGTGAGCATCCA 928
Qy 663 GTACGACAAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCAGGC 722
Db 929 GTACGACATAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCAGGC 988
Qy 723 AGCCCACTGCTTCAGGAAACATACCCGATGTGTTCACTGGAAGTGGGCGAGGCTCAGA 782
Db 989 AGCCCACTGCTTCAGGAAACATACCCGATGTGTTCACTGGAAGTGGGCGAGGCTCAGA 1048
Qy 783 CAAACTGGGAGCTTCCCATCCCTGGCTGGGCAAGATCATCATATTGAATTCACCCC 842
Db 1049 CAAACTGGGAGCTTCCCATCCCTGGCTGGGCAAGATCATCATATTGAATTCACCCC 1108
Qy 843 CATGTACCCCAAGACATGACATGCGCCCTCATGAAGTGCAGTTCCTCCTCTCTCTC 902
Db 1109 CATGTACCCCAAGACATGACATGCGCCCTCATGAAGTGCAGTTCCTCCTCTCTCTC 1168
Qy 903 AGGCACAGTCAGGCCCATCTGTCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 962
Db 1169 AGGCACAGTCAGGCTCATCTGTCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1228
Qy 963 ACTCTGGATCATTGGATGGGCTTTTACGAAGAGAAATGGAGGAGATGTTCTGACATACT 1022
Db 1229 ACTCTGGATCATTGGATGGGCTTTTACGAAGAGAAATGGAGGAGATGTTCTGACATACT 1288
Qy 1023 GCTGCGGCTCAGTCCAGTCTATTGACAGCACACGCTGCAATGCGAGCATCGGTACCA 1082
Db 1289 GCTGCGGCTCAGTCCAGTCTATTGACAGCACACGCTGCAATGCGAGCATCGGTACCA 1348
Qy 1083 GGGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCTCCCGAAGGGGGTGTGACACCTG 1142
Db 1349 GGGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCTCCCGAAGGGGGTGTGACACCTG 1408
Qy 1143 CGAGGTGACAGTGTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1202
Db 1409 CGAGGTGACAGTGTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1468
Qy 1203 CGTTAGTGGGCTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTC 1262
Db 1469 CGTTAGTGGGCTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTC 1328
Qy 1263 AGCCTATCTCAACTGGATCTCAATGTCTGGAAAGGCTGAGCTG 1305
Db 1529 AGCCTATCTCAACTGGATCTCAATGTCTGGAAAGGCTGAGCTG 1571

Query Match
; LENGTH: 2165
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-634

Query Match 99.2%; Score 1294; DB 15; Length 2165;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGATCTGACAGTATCAACCTCTGAACAGCCTCGATGTCGAAACCC 60
DB 310 ATGGATCTGACAGTATCAACCTCTGAACAGCCTCGATGTCGAAACCC 369
QY 61 CGTATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 370 CGTATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 429
QY 121 CTGGCGAGTATCATATTGTGTTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 180
DB 430 CTGGCGAGTATCATATTGTGTTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 489
QY 181 CTCTGGCGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGGAGAGCTGGAC 240
DB 490 CTCTGGCGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGGAGAGCTGGAC 549
QY 241 TGTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 300
DB 550 TGTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 609
QY 301 CGAGTCCGCTCTCCAAAGACCGATCCACATCTGAGGTGCTGGATCTGGCCACAGGGAAC 360
DB 610 CGAGTCCGCTCTCCAAAGACCGATCCACATCTGAGGTGCTGGATCTGGCCACAGGGAAC 669
QY 361 TGGTCTCTGCTGTTTTCAGCACTTCACAGAGCTCTCGCTGACAGCCTGTAGGAG 420
DB 670 TGGTCTCTGCTGTTTTCAGCACTTCACAGAGCTCTCGCTGACAGCCTGTAGGAG 729
QY 421 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCTG 480
DB 730 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCTG 789
QY 481 GATGTTGTTGAATACAGAAACAGCCAGGA-GCTTCGATGCGGAATCAAGTGGGCC 539
DB 790 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCC 849
QY 540 CTGCTCTCAGGCTCCCTGCTCTCCCTGACCTGTCTGCTGTGGAAAGAGCCTGAAGAC 599
DB 850 CTGCTCTCAGGCTCCCTGCTCTCCCTGACCTGTCTGCTGTGGAAAGAGCCTGAAGAC 909
QY 600 CCCCCTGTGGTGGGAGGAGGCTCTGTGATTTGGCTTGGCAGTCAAGT 659
DB 910 CCCCCTGTGGTGGGAGGAGGCTCTGTGATTTGGCTTGGCAGTCAAGT 969
QY 660 CCAATACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGCTCTCAC 719
DB 970 CCAATACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGCTCTCAC 1029
QY 720 GGCAGCCCACTGCTTCAGAAACATACCGATGTGTTCAATCGAAAGGTGGGAGGCTC 779
DB 1030 GGCAGCCCACTGCTTCAGAAACATACCGATGTGTTCAATCGAAAGGTGGGAGGCTC 1089
QY 780 AGCAAACTGGGAGCTTCCCACTCCCTGCTGTGGCCAGAGATCATCATTAATTCAA 839
DB 1090 AGCAAACTGGGAGCTTCCCACTCCCTGCTGTGGCCAGAGATCATCATTAATTCAA 1149
QY 840 CCCCATTATCCCAAGACATGACATCGCCCTCATGAAGCTGCAAGTCCCACTCACTTT 899
DB 1150 CCCCATTATCCCAAGACATGACATCGCCCTCATGAAGCTGCAAGTCCCACTCACTTT 1209
QY 900 CTCAGGACAGTCAAGCCCACTGCTGCTGCTCTTGTGATGAGGAGCTCACTCCAGCCAC 959
DB 1210 CTCAGGACAGTCAAGCCCACTGCTGCTGCTCTTGTGATGAGGAGCTCACTCCAGCCAC 1269

RESULT 15

US-09-851-588-5
; Sequence 5, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:

; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND N
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-851-588-5

Query Match 99.0%; Score 1292; DB 9; Length 2079;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCTCTGACAGTATCAACCTCTGAAAGAGCTTCGATGTCGAAACCCCTGCGAAACCCCG 62
DB 216 GGATCTCTGACAGTATCAACCTCTGAAAGAGCTTCGATGTCGAAACCCCTGCGAAACCCCG 275
QY 63 TATCCCATCGAGACCTTCAGAAAG-GTGGGATCCCATCATCATAGCACTACTGAGCC 121
DB 276 TATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY 122 TGGCGAGTATCATCAATTGTGTTGCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 181
DB 336 TGGCGAGTATCATCAATTGTGTTGCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 395
QY 182 TCTGCGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGGAGAGCTGACT 241
DB 396 TCTGCGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGAGGAGAGCTGACT 455
QY 242 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCAAGTG 301
DB 456 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCAAGTG 515

QY 302 CAGTCCGCTCTCCAGAGCCGATCCACACTGCGAGTCTGAGTCTGGCCACAGGGAAC 361
Db |||||
QY 516 CAGTCCGCTCTCCAGAGCCGATCCACACTGCGAGTCTGAGTCTGGCCACAGGGAAC 575
Db |||||
QY 362 GGTTCCTCTGCTGTTTTCAGAACTTCAGAGCTCTCGCTGAGACAGCTGTAGGCAGA 421
Db |||||
QY 576 GGTTCCTGCTGTTTTCAGAACTTCAGAACTTCAGAGCTCTCGCTGAGACAGCTGTAGGCAGA 635
Db |||||
QY 422 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGG 481
Db |||||
QY 636 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGG 695
Db |||||
QY 482 ATGTTGTTGAATACAGAAACAGCCAGAGCTTCGATCGGAACTCAAGTGGGCCCT 541
Db |||||
QY 696 ATGTTGTTGAATACAGAAACAGCCAGAGCTTCGATCGGAACTCAAGTGGGCCCT 755
Db |||||
QY 542 GTCTCTCAGGCTCCCTGCTGCTCCCTGCACTGCTTGTGCTGCTGGAAGAGCTGAAGACC 601
Db |||||
QY 756 GTCTCTCAGGCTCCCTGCTGCTCCCTGCACTGCTTGTGCTGCTGGAAGAGCTGAAGACC 815
Db |||||
QY 602 CCCGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTTCAGCATCC 661
Db |||||
QY 816 CCCGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTTCAGCATCC 875
Db |||||
QY 662 AGTAGCAAAACAGCAGCTCTGTGAGGAGGATCTGGAACCCCACTGGGTCTCTCAGG 721
Db |||||
QY 876 AGTAGCAAAACAGCAGCTCTGTGAGGAGGATCTGGAACCCCACTGGGTCTCTCAGG 935
Db |||||
QY 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGCTCAG 781
Db |||||
QY 936 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTGGGGCAGGCTCAG 995
Db |||||
QY 782 ACAACTGGCAGCTTCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
Db |||||
QY 996 ACAACTGGCAGCTTCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
Db |||||
QY 842 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAACTGCAGTTCCTCACTCTCTCT 901
Db |||||
QY 1056 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAACTGCAGTTCCTCACTCTCTCT 1115
Db |||||
QY 902 CAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGAGCTCACTCCAGCCACC 961
Db |||||
QY 1116 CAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGAGCTCACTCCAGCCACC 1175
Db |||||
QY 962 CACTCTGATCATTTGGATGGGCTTTTCAAGACAGAAATGGAGGAGAGATGCTGACATAC 1021
Db |||||
QY 1176 CACTCTGATCATTTGGATGGGCTTTTCAAGACAGAAATGGAGGAGAGATGCTGACATAC 1235
Db |||||
QY 1022 TGCTCAGGCTCAGTCCAGGTCAATTGACACACACGCTGCAATGCAAGCAGTCCGTACC 1081
Db |||||
QY 1236 TGCTCAGGCTCAGTCCAGGTCAATTGACACACACGCTGCAATGCAAGCAGTCCGTACC 1295
Db |||||
QY 1082 AGGGGAAAGTCACGAGAAAGATGATGTGAGGATCCCGAAGGGGTGTGACACCT 1141
Db |||||
QY 1296 AGGGGAAAGTCACGAGAAAGATGATGTGAGGATCCCGAAGGGGTGTGACACCT 1355
Db |||||
QY 1142 GCCAGGTCAGATGGTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1201
Db |||||
QY 1356 GCCAGGTCAGATGGTGGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1415
Db |||||
QY 1202 TCGTTAGCTGGGCTATGGCTGGGGGCTGAGGACACCCAGGAGTATACCAAGGTCT 1261
Db |||||
QY 1416 TCGTTAGCTGGGCTATGGCTGGGGGCTGAGGACACCCAGGAGTATACCAAGGTCT 1475
Db |||||
QY 1262 CAGCCTATCTCAACTGGATCTACATGCTGGAAGGCTGAGCTG 1305
Db |||||
QY 1476 CAGCCTATCTCAACTGGATCTACATGCTGGAAGGCTGAGCTG 1519
Db |||||

Search completed: June 8, 2004, 03:57:26
Job time : 607 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 59 Seconds
(without alignments)
2083.189 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQFLNSLDVKKPLRPK.....VYTKVSAYLNWYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_24Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2342	100.0	435	4	AAY72558 Human ser
2	2342	100.0	435	5	AG78577 Human ser
3	2342	100.0	435	6	ABU04931 Human exp
4	2342	100.0	435	6	ABU04912 Human exp
5	2338	99.8	435	2	AAY06437 Human pro
6	2338	99.8	435	6	ABU04930 Human exp
7	2337	99.8	437	5	AAG79359 CJA8 pref
8	2337	99.8	437	6	ABU04920 Human exp
9	2337	99.8	437	6	ABU04935 Human exp
10	2337	99.8	437	6	ABU04936 Human exp
11	2337	99.8	437	6	ABU04915 Human exp
12	2337	99.8	437	7	ADB80525 Ovarian c
13	2329	99.4	492	4	AAY72559 Human ser
14	2329	99.4	492	6	ABU04913 Human exp
15	2329	99.4	492	6	ABU04932 Human exp
16	2319	99.0	437	4	AAE06931 Human mem
17	2319	99.0	437	6	ABU04917 Human exp
18	2319	99.0	437	6	ABU04922 Human exp
19	2297.5	98.1	432	3	AAY59417 Human PRO
20	2297.5	98.1	432	4	AA566166 Protein o
21	2297.5	98.1	432	4	AAU29188 Human PRO
22	2297.5	98.1	432	4	AA387581 Human PRO
23	2297.5	98.1	432	5	ABG95906 Human sec
24	2297.5	98.1	432	5	ABU76535 Tumour-as
25	2297.5	98.1	432	6	ABU58564 Human PRO

26	2297.5	98.1	432	6	ABU88112	Novel hum
27	2297.5	98.1	432	6	ABU84427	Human sec
28	2297.5	98.1	432	6	ABR66301	Human sec
29	2297.5	98.1	432	6	ABR65691	Human sec
30	2297.5	98.1	432	6	ABU99631	Human sec
31	2297.5	98.1	432	6	ABU82870	Human PRO
32	2297.5	98.1	432	6	ABU89991	Novel hum
33	2297.5	98.1	432	6	ABR68240	Human sec
34	2297.5	98.1	432	6	ABU96293	Novel hum
35	2297.5	98.1	432	6	ABU92724	Human sec
36	2297.5	98.1	432	6	ABO08801	Human sec
37	2297.5	98.1	432	6	ABO02853	Human sec
38	2297.5	98.1	432	6	ABR75007	Human sec
39	2297.5	98.1	432	6	ABR94769	Human sec
40	2297.5	98.1	432	6	ABU85742	Human PRO
41	2297.5	98.1	432	6	ABU98902	Novel hum
42	2297.5	98.1	432	6	ABU98117	Novel hum
43	2297.5	98.1	432	6	ABU91823	Novel hum
44	2297.5	98.1	432	6	ABU89516	Human PRO
45	2297.5	98.1	432	6	ABU86357	Human sec

ALIGNMENTS

RESULT 1
AAY72558
ID AAY72558 standard; protein; 435 AA.
XX
AC AAY72558;
XX
XX
DT 02-MAY-2001 (first entry)
XX
DE Human seripancrin protein.
XX
KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
KW vaccine; cytostatic; cerebroprotective; vulnary; osteopathic.
XX
OS Homo sapiens.
XX
FN WO200104141-A2.
XX
PD 18-JAN-2001.
XX
PF 04-JUL-2000; 2000WO-EP006211.
XX
PR 12-JUL-1999; 99EP-00113428.
XX
PA (MERE) MERCK PATENT GMBH.
XX
FI Suendermann B, Hofmann U, Matzku S, Wilbert O;
XX
DR WPI; 2001-147177/15.
XX
N-PSDB; AAD02556.
XX
PT New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound
PT healing.
XX
PS Claim 2; Page 39-40; 45pp; English.
XX
CC The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine protease
CC family. This protein contains a transmembrane domain, a low density
CC lipoprotein (LDL) domain, protease domain and a scavenger receptor
CC cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the
CC specificity of seripancrin's intra and intermolecular interactions. The
CC polynucleotides and polypeptides of the invention are useful for treating
CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary

CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
CC diseases. Seripancrin genes are useful in chromosome localisation
CC studies, as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists and
CC antagonists useful for treating conditions associated with seripancrin
CC imbalance. These polypeptides are also useful as vaccines. The present
CC sequence is seripancrin protein. The seripancrin gene is located on human
CC chromosome 11q22-q23
XX
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 4; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKIVLDKYFF 60
Db 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKIVLDKYFF 60
QY 61 LCGQPLHFTPRKQLCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
Db 61 LCGQPLHFTPRKQLCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNSSGP 180
Db 121 WFSACFDNFTALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNSSGP 180
QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240
Db 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240
QY 241 AAHCFRKHDTDVNWKVRAGSKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQFPLTF 300
Db 241 AAHCFRKHDTDVNWKVRAGSKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQFPLTF 300
QY 301 SGTVPICLPFPDELTATPLIITGFTKQNGKMSDILLQASVQVIDSTRCNADAY 360
Db 301 SGTVPICLPFPDELTATPLIITGFTKQNGKMSDILLQASVQVIDSTRCNADAY 360
QY 361 QGEVTERKMCAGIPGGVDTCQDGGPLMVQSDQVHVGVSVGCGGSGTGVYTKV 420
Db 361 QGEVTERKMCAGIPGGVDTCQDGGPLMVQSDQVHVGVSVGCGGSGTGVYTKV 420
QY 421 SAYLWNIYNVWKAEL 435
Db 421 SAYLWNIYNVWKAEL 435

RESULT 2
AAG78577
ID AAG78577 standard; protein; 435 AA.

XX AAG78577;
XX
XX
DT 07-MAY-2002 (first entry)
XX
XX Human serine protease D-G amino acid sequence.

DE Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW anti-inflammatory; dermatological; anticoagulation; cancer; skin disorder;
KW neuropathic pain; inflammatory disorder; coagulation diathesis;
KW thrombosis; laundry detergent; skin care; gene therapy.
XX
OS Homo sapiens.
XX Key Location/Qualifiers
FH Region 31..52
FT /note= "hydrophobic transmembrane domain"
FT Active-site 202..203
FT /note= "residues spanning the zymogen cleavage site"
FT Active-site 243
FT /note= "catalytic triad residue"

FT Active-site 339 /note= "catalytic triad residue"
FT Misc-difference 361 /note= "Encoded by CTG"
FT Active-site 385 /note= "catalytic triad residue"
XX
FN W0200202011-AL.
XX
XX 10-JAN-2002.
XX
XX 08-JUN-2001; 2001WO-US018568.
XX
XX 30-JUN-2000; 2000US-00607745.
XX
XX (ORTH) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow AL, Qi J, Andrade-Gordon P;
XX
XX MPI: 2002-106601/14.
XX N-PSDB; AA164284.
XX
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin disorders,
XX or neuropathic pain.
XX
XX Claim 13; Fig 1B; 81pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, anti-inflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,
XX skin disorders, neuropathic pain, inflammatory disorders, or coagulation
XX diathesis/thrombosis. The polynucleotide encoding the protease is useful
XX for identifying, detecting or isolating mutant forms of DNA molecules
XX encoding the protease. The protease is useful for identifying modulators
XX of the functional protease. The D-G protein can be used for formulation
XX of compositions for laundry detergents and skin care products. Protease D
XX -G gene therapy may be used to introduce protease D-G into the cells of
XX target organisms. As the D-G protein is derived from a human, it is less
XX likely to produce an allergic reaction in sensitive individuals when used
XX in formulations for laundry detergents and skin care products. The
XX current sequence represents the human serine protease D-G amino acid
XX sequence
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 5; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKIVLDKYFF 60
Db 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKIVLDKYFF 60
QY 61 LCGQPLHFTPRKQLCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
Db 61 LCGQPLHFTPRKQLCGELDCPLGEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNSSGP 180
Db 121 WFSACFDNFTALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNSSGP 180
QY 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240
Db 181 CLSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 240

QY 241 AAHCFRKHDTVFNWVKVAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
Db 241 AAHCFRKHDTVFNWVKVAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDELTPTATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
Db 301 SGTVRPICLPFFDELTPTATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
QY 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVISWGYCGGSPSTPGYVTKV 420
Db 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVISWGYCGGSPSTPGYVTKV 420
QY 421 SAYLNWIYVNWKAEL 435
Db 421 SAYLNWIYVNWKAEL 435
RESULT 3
ID ABU04931 standard; protein; 435 AA.
XX AC ABU04931;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1597.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX FN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.
XX PR 04-DEC-2001; 2001US-0336780P.
XX PR 20-FEB-2002; 2002US-0358985P.
XX PA (ZYCO-) ZYCOS INC.
XX PI Chicx RM, Tomlinson AJ, Urban RG;
XX DR WPI; 2003-040607/03.
XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX PS Example 2; SEQ ID NO 1597; 134pp; English.
XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, or protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX
SQ Sequence 435 AA;
Query Match 100.0%; Score 2342; DB 6; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Mismatches 0; Gaps 0;
Matches 435; Conservative 0; Indels 0;
QY 1 MDPDSQPLNSLDVXPLRKPRIPIETFRKVGPIIIALLSLASIIIVVVLKIVLDKYYF 60
Db 1 MDPDSQPLNSLDVXPLRKPRIPIETFRKVGPIIIALLSLASIIIVVVLKIVLDKYYF 60
QY 61 LCGQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFEPGPAVAVRLSKDSTLOVLSATGN 120
Db 61 LCGQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFEPGPAVAVRLSKDSTLOVLSATGN 120
QY 121 WFSACFDNFEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGP 180
Db 121 WFSACFDNFEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEASVDSWPQVSIQYDKQHVCCGSILDPHWILT 240
Db 181 CLSGSLVSLHCLACGSKLTPRVVGGEASVDSWPQVSIQYDKQHVCCGSILDPHWILT 240
QY 241 AAHCFRKHDTVFNWVKVAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
Db 241 AAHCFRKHDTVFNWVKVAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDELTPTATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
Db 301 SGTVRPICLPFFDELTPTATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
QY 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVISWGYCGGSPSTPGYVTKV 420
Db 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVISWGYCGGSPSTPGYVTKV 420
QY 421 SAYLNWIYVNWKAEL 435
Db 421 SAYLNWIYVNWKAEL 435
RESULT 4
ID ABU04912 standard; protein; 435 AA.
XX AC ABU04912;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1578.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX FN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.
 PR 20-FEB-2002; 2002US-0358985P.
 PA (ZYCO-) ZYCOS INC.
 XX Chiciz RM, Tomlinson AJ, Urban RG;
 XX WPI; 2003-040607/03.
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.
 XX
 XX Example 2; SEQ ID NO 1578; 134pp; English.
 XX The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: This sequence does not appear in the printed from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 435 AA;
 SQ
 Query Match 100.0%; Score 2342; DB 6; Length 435;
 Best Local Similarity 100.0%; Pred. No. 2.8e-202; Mismatches 0; Gaps 0;
 Matches 435; Conservative 0; Indels 0;
 QY 1 MDPSDQPLNSLDVPLKRPIMPETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYF 60
 DB 1 MDPSDQPLNSLDVPLKRPIMPETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYF 60
 QY 61 LCGQLHPIPRKQCDGLDCLGEDEBHCVKSPPEGPAVAVRLSKORSTLQVILDSATGN 120
 DB 61 LCGQLHPIPRKQCDGLDCLGEDEBHCVKSPPEGPAVAVRLSKORSTLQVILDSATGN 120
 QY 121 WFSACFDNFTALAEATACQMGYSKPTFRAVEIGPDQDLVDVITENSQELRMENSSGP 180
 DB 121 WFSACFDNFTALAEATACQMGYSKPTFRAVEIGPDQDLVDVITENSQELRMENSSGP 180
 QY 181 CLSGSLVSLHCLACKSKLTPRVVGGEASVDSPFWQVSIQYDKHQVCGSILDPHVLVT 240
 DB 181 CLSGSLVSLHCLACKSKLTPRVVGGEASVDSPFWQVSIQYDKHQVCGSILDPHVLVT 240
 QY 241 AAHCFKHTDVFNVKVRAGSKLGFPSLAVAKIIITFENPMYKNDIALMKLQFLTF 300
 DB 241 AAHCFKHTDVFNVKVRAGSKLGFPSLAVAKIIITFENPMYKNDIALMKLQFLTF 300
 QY 301 SGTVRPILCLPFFDELTATPLWIIGWFTKQNGKMSDILLQASVQVVDSTRCNADAY 360
 DB 301 SGTVRPILCLPFFDELTATPLWIIGWFTKQNGKMSDILLQASVQVVDSTRCNADAY 360
 QY 361 QGEVTEKMKACIPGGVDTCGSGGGLMTYQSDQHWVGVSVGCGGPGSTPGVTVKV 420
 DB 361 QGEVTEKMKACIPGGVDTCGSGGGLMTYQSDQHWVGVSVGCGGPGSTPGVTVKV 420
 QY 421 SAYLNWIYNVWKAEI 435
 DB 421 SAYLNWIYNVWKAEI 435

RESULT 5
 AAY06437
 ID AAY06437 standard; protein; 435 AA.
 XX
 AC AAY06437;
 XX 27-SEP-1999 (first entry)
 DT XX
 XX Human protease HUPM-6.
 DE XX
 XX Serine protease; human; HUPM-6; cell proliferation; cancer;
 KW immune disorder; inflammation; therapy.
 XX
 OS Homo sapiens.
 XX
 PH Key
 FT Modified-site 26
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 59
 FT /note= "tyrosine kinase phosphorylation site"
 FT Modified-site 93
 FT /note= "casein kinase II phosphorylation site"
 FT Modified-site 128
 FT /note= "N-glycosylated"
 FT Modified-site 144
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 148
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 176
 FT /note= "N-glycosylated"
 FT Modified-site 137
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 200
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 231
 FT /note= "casein kinase II phosphorylation site"
 FT Active-site 243
 FT /note= "characteristic of serine protease"
 FT Modified-site 249
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT Modified-site 260
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 303
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 351
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 360
 FT /note= "tyrosine kinase phosphorylation site"
 FT Modified-site 365
 FT /note= "protein kinase C phosphorylation site"
 FT Active-site 385
 FT /note= "characteristic of serine protease"
 XX
 XX W09936550-A2.
 XX
 XX 22-JUL-1999.
 XX
 XX 12-JAN-1999; 99WO-US000655.
 XX
 XX 16-JAN-1998; 98US-00008271.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Bandman O, Hillman JL, Yue H, Guegler KJ, Corley NC, Tang YT;
 PI Shah P;
 PI WPI; 1999-430616/36.
 DR N-PSDB; AAX87154.
 XX
 XX Novel human protease molecules useful in the treatment of developmental
 PT disorders and/or cancers.
 XX

PS Claim 1; Page 74-75; 90pp; English.

XX The present sequence represents novel human protease HUPM-6, as deduced
CC from the consensus sequence (see AAX87154) of overlapping cDNA clones
CC obtained from various libraries. Northern analysis shows expression of
CC HUPM-6 in gastrointestinal, and male and female reproductive cDNA
CC libraries. Approximately 65% of these libraries are associated with
CC neoplastic disorders, and 22% with the immune response. The invention
CC provides 12 new human proteases, i.e. HUPM-1 to -12 (see AAY06432-43),
CC and the polynucleotides encoding them (see AAX87149-60). Also provided
CC are vectors, host cells and methods for producing HUPM polypeptides, as
CC well as agonists and antagonists of HUPM. Methods for treating or
CC preventing cell proliferative disorders and immune disorders using HUPM
CC or HUPM antagonists are claimed
XX
XX

SQ Sequence 435 AA;

Query Match 99.8%; Score 2338; DB 2; Length 435;
Best Local Similarity 99.8%; Pred. No. 6.5e-202;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYF 60
DB 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYF 60
QY 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPPEGPAVAVRLSKDRSTLQVLD SATGN 120
DB 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPPEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNFTALAEATACROMGYSSKPTFRVEIGPDQDLVDVEITENSQELMRNSSGP 180
DB 121 WFSACFDNFTALAEATACROMGYSSKPTFRVEIGPDQDLVDVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWLT 240
DB 181 CLSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWLT 240
QY 241 AAHCFRKHTEVFNKVRAGSKDLGSPFSLAVAKIIIEFNPMYKNDIALMKLOPLTF 300
DB 241 AAHCFRKHTEVFNKVRAGSKDLGSPFSLAVAKIIIEFNPMYKNDIALMKLOPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMWKACGIPGGVDTCQDGGPLMYQSDQVHVGVISWGYGCGGSPFGVYTKV 420
DB 361 QGEVTEKMWKACGIPGGVDTCQDGGPLMYQSDQVHVGVISWGYGCGGSPFGVYTKV 420
QY 421 SAYLNWYNWKAEL 435
DB 421 SAYLNWYNWKAEL 435

RESULT 6

ABU04930
ID ABU04930 standard; protein; 435 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human expressed protein tag (EPT) #1596.

Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
protease; protease inhibitor; transporter; cytoskeletal protein;
receptor; transcription factor; cancer; MHC;
major histocompatibility complex; myeloma; colon cancer; gastric cancer;
adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

Homo sapiens.

WO200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX 28-MAR-2001; 2001US-0279495P.

XX 21-MAY-2001; 2001US-0292544P.

XX 08-AUG-2001; 2001US-0310801P.

XX 01-OCT-2001; 2001US-0326370P.

XX 01-DEC-2001; 2001US-0336780P.

XX 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCOS INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX cytoskeletal proteins, receptors or transcription factors), useful for

XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

XX leukemia.

XX Example 2; SEQ ID NO 1596; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a

XX fragment of a kinase, phosphatase, protease, protease inhibitor,

XX transporter, cytoskeletal protein, receptor or transcription factor. The

XX polypeptide is useful as an immunogenic composition for eliciting in a

XX mammal an immunogenic response directed against any of the purified

XX polypeptide. The purified polypeptide, or the antibody that binds to this

XX polypeptide, is useful for treating cancer. The polypeptide is also

XX useful for identifying compounds that binds to a naturally processed

XX class I or class II MHC-binding polypeptide. The polypeptides and

XX polynucleotides are particularly useful for treating or preventing

XX lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,

XX lymphoma or leukaemia. These are also useful for screening agents for

XX treating the above mentioned diseases. This sequence represents an

XX expressed protein tag (EPT) isolated from human tissue for translational

XX profiling. Note: this sequence does not appear in the printed

XX specification but was obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 435 AA;

Query Match 99.8%; Score 2338; DB 6; Length 435;

Best Local Similarity 99.8%; Pred. No. 6.5e-202;

Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYF 60

DB 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKYF 60

QY 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPPEGPAVAVRLSKDRSTLQVLD SATGN 120

DB 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPPEGPAVAVRLSKDRSTLQVLD SATGN 120

QY 121 WFSACFDNFTALAEATACROMGYSSKPTFRVEIGPDQDLVDVEITENSQELMRNSSGP 180

DB 121 WFSACFDNFTALAEATACROMGYSSKPTFRVEIGPDQDLVDVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWLT 240

DB 181 CLSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWLT 240

QY 241 AAHCFRKHTEVFNKVRAGSKDLGSPFSLAVAKIIIEFNPMYKNDIALMKLOPLTF 300

DB 241 AAHCFRKHTEVFNKVRAGSKDLGSPFSLAVAKIIIEFNPMYKNDIALMKLOPLTF 300

QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360

DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAY 360

QY 361 QGEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKV 420
 DB 361 QGEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKV 420
 QY 421 SAYLNWIYNVWKAEL 435
 DB 421 SAYLNWIYNVWKAEL 435

RESULT 7

AAG79359
 ID AAG79359 standard; protein; 437 AA.

AC AAG79359;

XX 21-AUG-2002 (first entry)

XX CJA8 preferred sequence.

XX Colorectal cancer; CGA7; CJA8; modulating protein; screening;

KW drug candidate; vaccine.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 36..63

FT /notes="Transmembrane domain"

XX US2002042067-A1.

XX 11-APR-2002.

XX 08-MAY-2001; 2001US-00851588.

XX 17-AUG-2000; 2000US-00642252.

PR 06-SEP-2000; 2000US-00656002.

XX (MACK/) MACK D.

PA (GISH/) GISH K C.

PA (WILS/) WILSON K E.

XX Mack D, Gish KC, Wilson KE;

XX WPI: 2002-453647/48.

DR N-PSDB; AA172976.

XX Screening drug candidates for treating colorectal cancer, comprises
 PT determining the effect of the candidate on the expression profile gene of
 PT CGA7 or CJA8.

XX Disclosure; Fig 8; 40pp; English.

XX The sequences given in AAG79356-59 show the colorectal cancer proteins,
 CC CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer
 CC modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11
 CC (CJA8). These sequences may be used in the method of the invention for
 CC screening drug candidates. The method comprises adding a drug candidate
 CC to a cell that expresses an expression profile gene encoding CGA7, CJA8
 CC or fragments and determining the effect of the drug candidate on the
 CC expression of the expression profile gene. The new methods are used to
 CC screen bioactive agents for the ability to bind to or modulate the
 CC activity of CGA7 or CJA8 and evaluate the effect of a candidate
 CC colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit the
 CC activity of CGA7 or CJA8, respectively, and is used to screen for an
 CC agent that can interfere with the binding of CGA7 or CJA8 to the
 CC antibody. The antibody can be used to treat colorectal cancer. The
 CC antibody or a fragment of it is used to localize a therapeutic group to a
 CC colorectal cancer tissue, where the therapeutic group is a cytotoxic
 CC agent or a radioisotope. Antisense molecules are used to inhibit
 CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CJA8
 CC are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it are used
 CC to elicit an immune response. CGA7 or CJA8 is used to determine the
 CC prognosis of an individual with colorectal cancer. Nucleic acid encoding

CC CGA7 or CJA8 can be used in vaccines

XX Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 5; Length 437;

Best Local Similarity 100.0%; Pred. No. 86-202;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 61
 DB 4 DPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 63
 QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 DB 64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 181
 DB 124 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWFPQVSIQYDKQHVCGGSIILDPHWLTA 241
 DB 184 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWFPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHTDVFNKVRAGSDKLSGSPFSLAVAKIIIEFNPMPKNDIALMKLOFLTPS 301
 DB 244 AHCFRKHTDVFNKVRAGSDKLSGSPFSLAVAKIIIEFNPMPKNDIALMKLOFLTPS 303
 QY 302 GTVRPICLPFFDEELTPATPLWTIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 361
 DB 304 GTVRPICLPFFDEELTPATPLWTIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 363
 QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKVS 421
 DB 364 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKVS 423
 QY 422 AYLNWIYNVWKAEL 435
 DB 424 AYLNWIYNVWKAEL 437

RESULT 8

ABU04920

ID ABU04920 standard; protein; 437 AA.

XX AC ABU04920;

XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1596.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

XX WO200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

XX 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCOS INC.

XX PI Chiciz RM, Tomlinson AJ, Urban RG;
XX DR WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX
XX Example 2; SEQ ID NO 1586; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor. The
XX polypeptide is useful as an immunogenic composition for eliciting in a
XX mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to this
XX polypeptide, is useful for treating cancer. The polypeptide is also
XX useful for identifying compounds that binds to a naturally processed
XX class I or class II MHC-binding polypeptide. The polypeptides and
XX polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling. Note: This sequence does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 437 AA;
XX
XX Query Match 99.8%; Score 2337; DB 6; Length 437;
XX Best Local Similarity 100.0%; Pred. No. 8e-202;
XX Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 61
XX DB 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 63
XX
XX QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNW 121
XX DB 64 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNW 123
XX
XX QY 122 FSACPDNTEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGPC 181
XX DB 124 FSACPDNTEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGPC 183
XX
XX QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 241
XX DB 184 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
XX
XX QY 242 AHCERKHTDVFNVKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
XX DB 244 AHCERKHTDVFNVKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
XX
XX QY 302 GTVRPICI.PFDEELTRATPLMIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 361
XX DB 304 GTVRPICI.PFDEELTRATPLMIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 363
XX
XX QY 362 GEVTEKMWKAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 421
XX DB 364 GEVTEKMWKAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 423
XX
XX QY 422 AYLNWIYNWKAEL 435
XX DB 424 AYLNWIYNWKAEL 437
XX
XX RESULT 9
XX ID ABU04935 standard; protein; 437 AA.
XX

AC ABU04935;
XX
XX DT 29-JAN-2003 (first entry)
XX
XX DE Human expressed protein tag (EPT) #1601.
XX
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
XX protease; protease inhibitor; transporter; cytoskeletal protein;
XX receptor; transcription factor; cancer; MHC;
XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;
XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX
XX OS Homo sapiens.
XX
XX WO200278524-A2.
XX
XX PD 10-OCT-2002.
XX
XX PF 28-MAR-2002; 2002WO-US0009671.
XX
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.
XX PR 04-DEC-2001; 2001US-0336780P.
XX PR 20-FEB-2002; 2002US-0358985P.
XX
XX (ZYCO-) ZYCOS INC.
XX
XX PA Chiciz RM, Tomlinson AJ, Urban RG;
XX
XX PI WPI; 2003-040607/03.
XX
XX DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX
XX PS Example 2; SEQ ID NO 1601; 134pp; English.
XX
XX CC The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor. The
XX polypeptide is useful as an immunogenic composition for eliciting in a
XX mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to this
XX polypeptide, is useful for treating cancer. The polypeptide is also
XX useful for identifying compounds that binds to a naturally processed
XX class I or class II MHC-binding polypeptide. The polypeptides and
XX polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling. Note: This sequence does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 437 AA;
XX
XX Query Match 99.8%; Score 2337; DB 6; Length 437;
XX Best Local Similarity 100.0%; Pred. No. 8e-202;
XX Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 61
XX DB 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 63
XX
XX QY 62 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNW 121
XX DB 64 CGQPLHFIPRKQLCDGELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNW 123
XX
XX QY 122 FSACPDNTEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGPC 181
XX DB 124 FSACPDNTEALAEATACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGPC 183
XX
XX QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 241
XX DB 184 LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
XX
XX QY 242 AHCERKHTDVFNVKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
XX DB 244 AHCERKHTDVFNVKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
XX
XX QY 302 GTVRPICI.PFDEELTRATPLMIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 361
XX DB 304 GTVRPICI.PFDEELTRATPLMIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 363
XX
XX QY 362 GEVTEKMWKAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 421
XX DB 364 GEVTEKMWKAGIPEGGVDTCCQDSGGPLMYQSDQWHVVGIVSWGCGGPGSTPGVYTKVS 423
XX
XX QY 422 AYLNWIYNWKAEL 435
XX DB 424 AYLNWIYNWKAEL 437
XX
XX RESULT 9
XX ID ABU04935 standard; protein; 437 AA.
XX

Db 124 FSAACFDNFTEALAEATACRQMGVSSKPTFRAVEIGDPDQLDVEITENSQELRMENSSGPC 183
Qy 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
Db 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 243
Qy 242 AHCFRKHTDVFNNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCFRKHTDVFNNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
Qy 302 GTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVVDSTRCNADDAAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVVDSTRCNADDAAYQ 363
Qy 362 GEVTERKMCAGIPEGVDTCQDSDGGPLMYQSDQVHVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTERKMCAGIPEGVDTCQDSDGGPLMYQSDQVHVGIVSWGCGGSPSTPGVYTKVS 423
Qy 422 AYLNWIYNNWKAEL 435
Db 424 AYLNWIYNNWKAEL 437

RESULT 10
ABU04936
ID ABU04936 standard; protein; 437 AA.
XX AC ABU04936;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1602.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX PN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0328370P.
XX PR 04-DEC-2001; 2001US-0336780P.
XX PR 20-FEB-2002; 2002US-0358985P.
XX PA (ZYCO-) ZYCOS INC.
XX PI Chicx RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX PS Example 2; SEQ ID NO 1602; 134pp; English.
XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified

CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: this sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 6; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 DPDSQPLNSLDVPLKPRIPMETPRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 61
Db 4 DPDSQPLNSLDVPLKPRIPMETPRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 63
Qy 62 CGOPLHFIPIRKOICDGLDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLDATGNW 121
Db 64 CGOPLHFIPIRKOICDGLDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLDATGNW 123
Qy 122 FSACFDNFTEALAEATACRQMGVSSKPTFRAVEIGDPDQLDVEITENSQELRMENSSGPC 181
Db 124 FSACFDNFTEALAEATACRQMGVSSKPTFRAVEIGDPDQLDVEITENSQELRMENSSGPC 183
Qy 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
Db 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 243
Qy 242 AHCFRKHTDVFNNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCFRKHTDVFNNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
Qy 302 GTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVVDSTRCNADDAAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVVDSTRCNADDAAYQ 363
Qy 362 GEVTERKMCAGIPEGVDTCQDSDGGPLMYQSDQVHVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTERKMCAGIPEGVDTCQDSDGGPLMYQSDQVHVGIVSWGCGGSPSTPGVYTKVS 423
Qy 422 AYLNWIYNNWKAEL 435
Db 424 AYLNWIYNNWKAEL 437

RESULT 11
ABU04915
ID ABU04915 standard; protein; 437 AA.
XX AC ABU04915;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1581.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX PN WO200278524-A2.
XX PD 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.
XX 28-MAR-2001; 2001US-0279495P.
XX 21-MAY-2001; 2001US-0292544P.
XX 08-AUG-2001; 2001US-0310801P.
XX 01-OCT-2001; 2001US-0326370P.
XX 04-DEC-2001; 2001US-0336780P.
XX 20-FEB-2002; 2002US-0358985P.
XX (ZYCO-) ZYCOS INC.
XX Chicz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX Example 2; SEQ ID NO 1581; 134pp; English.
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor. The
XX polypeptide is useful as an immunogenic composition for eliciting in a
XX mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to this
XX polypeptide, is useful for treating cancer. The polypeptide is also
XX useful for identifying compounds that binds to a naturally processed
XX class I or class II MHC-binding polypeptide. The polypeptides and
XX polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling. Note: This sequence does not appear in the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 6; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVIDKYFL 61
DB 4 DPDSQPLNSLDVPLKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVIDKYFL 63
QY 62 CQQLPHFI-PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 121
DB 64 CQQLPHFI-PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123
QY 122 FSACFDNTEALAEACQMGYSKPTFAVEIGDPDQDLDVVEITENSQELMRNMSGPC 181
DB 124 FSACFDNTEALAEACQMGYSKPTFAVEIGDPDQDLDVVEITENSQELMRNMSGPC 183
QY 182 LSGSLVSLHCLACGSLKTPRVVGGEASVDSWPMQVSIQYDKQKVCGSLIDPHWVLT 241
DB 184 LSGSLVSLHCLACGSLKTPRVVGGEASVDSWPMQVSIQYDKQKVCGSLIDPHWVLT 243
QY 242 AHCFRKHTDVFNWKVRAGSKLGSPPSLAVAKIIIEFNPMYKNDIALMKLOPPLTFS 301
DB 244 AHCFRKHTDVFNWKVRAGSKLGSPPSLAVAKIIIEFNPMYKNDIALMKLOPPLTFS 303
QY 302 GTVRICLPFFDEELTPATLWIIQWGTTKONGKGVSDILLQASVOVLDSTRCNADDAAYQ 361
DB 304 GTVRICLPFFDEELTPATLWIIQWGTTKONGKGVSDILLQASVOVLDSTRCNADDAAYQ 363
QY 362 GEVTEKMWKAGIPEGGVDTCCGDSGGPLMYQSDQHHVGVISWVGCGGPGTGGVTKVS 421

DB 364 GEVTEKMWKAGIPEGGVDTCCGDSGGPLMYQSDQHHVGVISWVGCGGPGTGGVTKVS 423
QY 422 AYLNWIYNVWKAE 435
DB 424 AYLNWIYNVWKAE 437
RESULT 12
ADB80525 standard; protein; 437 AA.
XX ADB80525;
XX AC ADB80525;
XX DT 04-DEC-2003 (first entry)
XX DE Ovarian cancer-associated protein #45.
XX cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;
XX post-operative chemotherapy; radiation therapy; tumour prognosis;
XX pre-cancerous lesion detection.
XX OS Homo sapiens.
XX PN W02002102235-A2.
XX PD 27-DEC-2002.
XX PF 18-JUN-2002; 2002WO-US019297.
XX PR 18-JUN-2001; 2001US-0299234P.
XX PR 27-AUG-2001; 2001US-0315287P.
XX PR 05-SEP-2001; 2001US-0317544P.
XX PR 13-NOV-2001; 2001US-0350666P.
XX PR 12-APR-2002; 2002US-0372246P.
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX FI Mack DH, Gish KC;
XX WPI; 2003-167431/16.
XX N-PSDB; ADB80524.
XX Detecting an ovarian cancer-associated transcript in a cell from a
XX patient, comprises contacting a biological sample from the patient with a
XX polynucleotide that hybridizes to an ovarian cancer gene.
XX Claim 13; Page 305; 332pp; English.
XX The invention relates to a method of detecting an ovarian cancer-
XX associated transcript in a cell from a patient, by contacting a
XX biological sample from the patient with a polynucleotide that selectively
XX hybridizes to a sequence at least 80% identical to any of one of 80
XX nucleic acid sequences given in the specification. The method is useful
XX in diagnosing ovarian cancer and in identifying and using agents and/or
XX targets that inhibit ovarian cancer. The nucleic acid molecule,
XX polypeptide and the antibody may also be used in detecting ovarian
XX cancers, monitoring and early detection of relapse following treatment,
XX monitoring response to therapy, selecting patients for post-operative
XX chemotherapy or radiation therapy, in selecting mode of therapy,
XX determining tumour prognosis, early detection of pre-cancerous lesions,
XX and as vaccines. This sequence corresponds to one of the proteins used
XX for the detection method of the invention.
XX Sequence 437 AA;
Query Match 99.8%; Score 2337; DB 7; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVIDKYFL 61
DB 4 DPDSQPLNSLDVPLKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVIDKYFL 63

QY 62 CQQLHFIIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLDATGNW 121
 Db 64 CQQLHFIIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLDATGNW 123
 QY 122 FSACFDNFTFTEALAEACQMGYSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 181
 Db 124 FSACFDNFTFTEALAEACQMGYSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 183
 QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQDKHVCVCGSILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQDKHVCVCGSILDPHWLTA 243
 QY 242 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 301
 Db 244 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 303
 QY 302 GTVRPTICLPPFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVLDSTRCNADDAQ 361
 Db 304 GTVRPTICLPPFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVLDSTRCNADDAQ 363
 QY 362 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVVGIWSWYCGGPGSTPGVYTKV 421
 Db 364 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVVGIWSWYCGGPGSTPGVYTKV 423
 QY 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 13

AAV72559
 ID AAY72559 standard; protein; 492 AA.

XX AAY72559;

XX AAY72559;

XX 02-MAY-2001 (first entry)

XX Human seripancrin variant #1 protein.

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX arthritis; chronic obstructive pulmonary disorder; COPD; cancer;

XX osteoporosis; aberrant wound healing; angiogenesis; diabetes;

XX inflammatory disorder; stroke; cardiovascular disease; gene therapy;

XX vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic.

XX Homo sapiens.

XX WO200104141-A2.

XX 18-JAN-2001.

XX 04-JUL-2000; 2000WO-EP006211.

XX 12-JUL-1999; 99EP-00113428.

XX (MERE) MERCK PATENT GMBH.

XX Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX WPI; 2001-147177/15.

XX N-PSDB; AAD02557.

XX New extracellular serine protease Seripancrin, useful for treating

XX cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,

XX diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound

XX healing.

XX Disclosure; Page 43-44; 45pp; English.

XX The present invention relates to seripancrin polynucleotides, and

XX polypeptides encoded by them. Seripancrin are members of serine protease

XX family. This protein contains a transmembrane domain, a low density

XX lipoprotein (LDL) domain, protease domain and a scavenger receptor

CC cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the
 CC specificity of seripancrin's intra and intermolecular interactions. The
 CC polynucleotides and polypeptides of the invention are useful for treating
 CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary
 CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
 CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
 CC diseases. Seripancrin genes are useful in chromosome localisation
 CC studies, as tools for tissue expression studies and also in gene therapy.
 CC The polypeptides of the invention are used for identifying agonists and
 CC antagonists useful for treating conditions associated with seripancrin
 CC imbalance. These polypeptides are also useful as vaccines. The present
 CC sequence is seripancrin variant #1 protein. The seripancrin gene is
 CC located on human chromosome 11q22-q23
 XX
 SQ Sequence 492 AA;

Query Match 99.4%; Score 2329; DB 4; Length 492;
 Best Local Similarity 100.0%; Pred. No. 4.9e-201;
 Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVIDKYIF 60
 Db 1 MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVIDKYIF 60
 QY 61 LCGQPLHFIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLDATGN 120
 Db 61 LCGQPLHFIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLDATGN 120
 QY 121 WFSACFDNFTFTEALAEACQMGYSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 180
 Db 121 WFSACFDNFTFTEALAEACQMGYSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 180
 QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQDKHVCVCGSILDPHWLTA 240
 Db 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQDKHVCVCGSILDPHWLTA 240
 QY 241 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
 Db 241 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
 QY 301 SGTVRPTICLPPFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVLDSTRCNADDAQ 360
 Db 301 SGTVRPTICLPPFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVLDSTRCNADDAQ 360
 QY 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVVGIWSWYCGGPGSTPGVYTKV 420
 Db 361 QGEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQVHVVGIWSWYCGGPGSTPGVYTKV 420
 QY 421 SAYLNWIYNNWKAEL 432
 Db 421 SAYLNWIYNNWKAEL 432

RESULT 14

ABU04913
 ID ABU04913 standard; protein; 492 AA.

XX ABU04913;

XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1579.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

XX protease; protease inhibitor; transporter; cytoskeletal protein;

XX receptor; transcription factor; cancer; MHC;

XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;

XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX WO200278524-A2.

XX

Best Local Similarity 100.0%; Pred. No. 4.9e-201;			
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MDPDSQPLNSLDVKPLRKPRIPIIPIIALLSLASIIIVVLIKVIDKYIF	60
Db	1	MDPDSQPLNSLDVKPLRKPRIPIIPIIALLSLASIIIVVLIKVIDKYIF	60
Qy	61	LCGQPLHFIIPRQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKORSTLOVILDSATGN	120
Db	61	LCGQPLHFIIPRQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKORSTLOVILDSATGN	120
Qy	121	WFSACFDNFTALAEACRQGYSSKPTFRAVEIGPDODLDWEITENSQELRNRSSGP	180
Db	121	WFSACFDNFTALAEACRQGYSSKPTFRAVEIGPDODLDWEITENSQELRNRSSGP	180
Qy	181	CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLT	240
Db	181	CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLT	240
Qy	241	AAHCFRKHTDVFNWVRAGSKLGSFPISLAVAKIIIEFNPMYKNDNDIALMKLOEPLTF	300
Db	241	AAHCFRKHTDVFNWVRAGSKLGSFPISLAVAKIIIEFNPMYKNDNDIALMKLOEPLTF	300
Qy	301	SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAY	360
Db	301	SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAY	360
Qy	361	QGEVTEKXMCAGIPGGVDTCQDSGGPLMYQSDOWHVGVISWYCGGPGSTPGVYTKV	420
Db	361	QGEVTEKXMCAGIPGGVDTCQDSGGPLMYQSDOWHVGVISWYCGGPGSTPGVYTKV	420
Qy	421	SAYLNWIYVNVK	432
Db	421	SAYLNWIYVNVK	432

Search completed: June 7, 2004, 08:40:34
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 20 Seconds.
(without alignments)
2092.166 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWYNVWRAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	584.5	25.0	417	1	S00845	hepsin (EC 3.4.21.)
2	582.5	24.9	1019	1	A58318	enteropeptidase (E
3	571	24.4	1035	1	A43090	enteropeptidase (E
4	568.5	24.3	416	1	S33777	hepsin (EC 3.4.21.)
5	552	23.6	1034	1	A53663	enteropeptidase (E
6	511	21.8	638	1	KQSTPL	plasma kallikrein
7	508	21.7	638	1	KQMSPL	plasma kallikrein
8	505	21.6	436	2	JX0172	acrosin (EC 3.4.21
9	504.5	21.5	1524	2	T30337	polyprotein - Afri
10	503	21.5	418	2	A37344	acrosin (EC 3.4.21
11	501.5	21.4	437	2	S18407	acrosin (EC 3.4.21
12	493.5	21.1	421	2	S29599	acrosin (EC 3.4.21
13	491.5	21.0	638	1	KQHUP	plasma kallikrein
14	488.5	20.9	1113	2	JE0315	low-density lipopor
15	481	20.5	421	1	S11674	acrosin (EC 3.4.21
16	478.5	20.4	415	1	A34170	acrosin (EC 3.4.21
17	474	20.2	855	2	JC7731	membrane-bound arg
18	470	20.1	761	2	JCS759	brain-specific ser
19	468.5	20.0	431	2	S47538	acrosin (EC 3.4.21
20	461.5	19.7	277	2	S35340	trypsin (EC 3.4.21
21	460.5	19.7	263	1	KYFTB	chymotrypsin (EC 3
22	458.5	19.6	267	2	S40006	trypsin (EC 3.4.21
23	458.5	19.6	790	1	PLFG	plasmin (EC 3.4.21
24	457.5	19.5	625	1	KFHU1	coagulation factor
25	455.5	19.4	812	1	PLMS	plasmin (EC 3.4.21
26	449	19.2	420	2	A55283	acrosin (EC 3.4.21
27	448	19.1	810	2	I46260	plasmin (EC 3.4.21
28	447.5	19.1	343	1	A57014	proctasin (EC 3.4
29	445.5	19.0	455	2	A61545	plasmin (EC 3.4.21

ALIGNMENTS

RESULT 1

S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: S00845
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dome
A:Reference number: S00845; MUID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEY>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; P:ID:g32064
C:Genetics:
A:Gene: GDB:HPN; TMRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TMN>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 25.0%; Score 584.5; DB 1; Length 417;
Best Local Similarity 31.5%; Pred.No. 7.5e-41;
Matches 140; Conservative 60; Mismatches 161; Indels 83; Gaps 11;

QY	19	KPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLKDYFLCGQLHPKQLCDGE	78
DB	15	RPKVAULT---AGTLLELTAGASWAIVALLR-----	45
QY	79	LDCLGDEDEHCVKSPFEGPAVAVRLSKRSTLQVLDSATGNWFSACDFNFTALATAC	138
DB	46	-----SDQE-----FLYPQVQSSADARLMVFKEGTWRLLCSSRSNARVAGLSC	90
QY	139	ROMCYSKPTFRFAVEICPDQDLDWTEITN-----SQELRMNSSGPECLS	183
DB	91	EEMGF-----LRAL-----THSELVDRTAGANGTSGFCVDEGRLPHTORLLEVISVCDCPR	142
QY	184	GSLSVSLHCLACG-KSLKTPRVVGEEASVDSPWQVSIQYDKHVCVGSILDPHWLTA	242
DB	143	GRFLAALICQDCGRKLPVDRIVGGRTSLGRWPQVSLRYDGAHLCCGSLSGDWLTA	202
QY	243	HCPRKHTDVN-MKVRAGSKLGSFPPLAVAKIIIEFNMPY-----KNDIALMK	293
DB	203	HCFFPERNVLRLRWEVFAQAQSPHGLQGVQAVVHHGYLPFRDPNSENENDIALVH	262
QY	294	LQFPLTSGTVRPTCLPFDEELTPATPLMTIGWGFTKQNGKMSDILLOASVQVIDSTR	353
DB	263	LSSPLPTEVTIQVCLPAAGQALVDGKICVTGTGNT-QYVGGQAGVLQEARVPIISNDV	321

QY 354 CNADAYQGEVTERKMCAGIEGGVDTTCQDGGPLMTQ-----SDQHVHVGIVSWGYGC 408
 Db 322 CNGADFYGNKPKPCFAGPEGGIDACQDGGPFCVEDSISRTPRWRLCGIVSWGTGC 381
 QY 409 GPFSTPGVYTKVSAYLWVYVWK 432
 Db 382 ALAQKPGVYTKVSDPFWIFQAIK 405

RESULT 2
 A56318
 enteropeptidase (EC 3.4.21.9) precursor [validated] - human
 N;Alternate names: enterokinase
 C;Species: Homo sapiens (man)
 C;Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
 C;Accession: A56318; B43090
 R;Kitamoto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.
 Biochemistry 34, 4562-4568, 1995
 A;Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
 A;Reference number: A56318; MUID:59234679; PMID:7718557
 A;Accession: A56318
 A;Molecule type: mRNA
 A;Residues: 1-1019 <KIT>
 A;Cross-references: GB:U09860; NID:G746412; PIDN:AAC50138.1; PID:G746413
 R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A;Reference number: A43090; MUID:94329561; PMID:8052624
 A;Accession: B43090
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 749-1019 <KIT>
 A;Cross-references: GB:U09860
 C;Comment: The mechanism of association with the membrane of the intestinal brush border
 (located below) or with amino-terminal myristoylation of the heavy chain.
 C;Genetics:
 A;Gene: GDB:PRSS7
 A;Cross-references: GDB:384083; OMIM:226200
 A;Map position: 21q21-21q21
 C;Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
 ducts.
 C;Function:
 A;Description: cleaves activation peptide from trypsinogen to produce active trypsin
 A;Pathway: intestinal digestive hydrolase cascade
 C;Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep
 C;Keywords: Glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
 F;1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
 F;184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;342-504/Domain: MAM homology <MAM>
 F;526-631/Domain: Clr/Cls repeat homology <Clr>
 F;643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
 F;785-1019/Product: enteropeptidase light chain #status predicted <LCH>
 F;785-1014/Domain: trypsin homology <TRY>
 F;116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
 F;772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
 F;825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 24.9%; Score 582.5; DB 1; Length 1019;
 Best Local Similarity 33.8%; Pred. No. 3,3e-40;
 Matches 135; Conservative 62; Mismatches 132; Indels 71; Gaps 15;

QY 67 HF-----IPKQLGDLGELDCPLGDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSATG 119
 Db 647 HFQKNGECVPLWNLCDGHLHCEGSDGADCVRF-----NGRTN 686
 QY 120 N-----WFSNCFNFTFALAEACROMGY-----SSKPTFAVEIGPDQDLWVE 164
 Db 687 NGLVRFRIQSIWHTACAEWTTQISNDVCQLGLGSGNSSKPIF-STDGGPFVFKLNTAP 745

QY 165 -----ITNSOELMRNSSGPGCLSGSLVSLHC--LACGKSILK-----TPRVVGGEEASVDGM 214
 Db 746 DGHLLITPSSQ-----CLQDSILRLQCNHKSCKGLAAQDITPKIVGGSSNAKEGAW 796
 QY 215 PMQVSIQYDKHQHCGGSIIDPHWLVTAHC-FKFKHTDVFNWVKVRAGSDKLGGFPFSLAVAK 273
 Db 797 PMVGLYGGRLCGASLVSSDLVSAACHVGVGNLPSKWTAILGLHMKSNLTSPTVP 856
 QY 274 III--TEFNMY---PKNDIALMKIQLPFTSGTVRPICLFFDEBELTPATLWIGWG 328
 Db 857 RLIDEIVINPHYNNRRKNDIAMHLEFKVNYTDYIQPICLPEENQVFPFGNCSLAGWG 916
 QY 329 FTKQGGKMSDILQASVOVIDSTRCNADDAVQGETKMMKAGIEGGVDTTCQDGGSP 388
 Db 917 -TVYQGTNTANILQEDVPLLSNERCQ-QOMPYNITENWICAGYEGGIDSCQDGGSP 974
 QY 389 LMYQ-SDQHVHVGIVSWGYCGGPGSTPGVYTKVSAYLWV 427
 Db 975 LMCQENRWFLAGVTSFGYKCALPNRPGVYVSRFTW 1014

RESULT 3
 A43090
 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
 N;Alternate names: enterokinase
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Apr-2003
 C;Accession: A43090; A48874; A61436
 R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A;Reference number: A43090; MUID:94329561; PMID:8052624
 A;Accession: A43090
 A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1035 <KIT>
 A;Cross-references: GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411
 R;Experimental source: small intestine
 R;LaValle, E.R.; Rehmentulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; I
 J. Biol. Chem. 268, 23311-23317, 1993
 A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of b
 A;Reference number: A48874; MUID:94043122; PMID:8226855
 A;Accession: A48874
 A;Molecule type: mRNA
 A;Residues: 801-1035 <LAV>
 A;Cross-references: GB:U19663; NID:G416131; PIDN:AAA16035.1; PID:G416132
 A;Note: parts of this sequence, including the amino end of the mature protein, were confi
 R;Light, A.; Janska, H.
 J. Protein Chem. 10, 475-480, 1991
 A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
 A;Reference number: A61436; MUID:92189715; PMID:1799406
 A;Accession: A61436
 A;Molecule type: protein
 A;Residues: 801-807,'Y',809-827 <LIIG>
 C;Comment: The mechanism of association with the membrane of the intestinal brush border
 embrane attachment using a signal-anchor sequence.
 C;Complex: Conversion from membrane-bound to soluble forms may involve further processing
 C;Comment: mature enteropeptidase is variously reported to contain two (heavy and light)
 lide linked
 C;Function:
 A;Description: cleaves propeptide from trypsinogen to produce active trypsin
 A;Pathway: intestinal digestive hydrolase cascade
 C;Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep
 C;Keywords: Glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein;
 F;22-38/Domain: transmembrane #status predicted <TM>
 F;52-117/Product: enteropeptidase mini chain #status predicted <MCH>
 F;118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
 F;199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;358-520/Domain: MAM homology <MAM>
 F;542-647/Domain: Clr/Cls repeat homology <Clr>
 F;659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
 F;801-1035/Product: enteropeptidase light chain #status predicted <LCH>

F:801-1030/Domain: trypsin homology <TRY>
F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
F:788-912,826-942,926-993,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.4%; Score 571; DB 1; Length 1035;
Best Local Similarity 35.0%; Pred. No. 3e-39;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

QY 69 IPRKLCDELCDPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
Db 672 IPLVNLCDGPFCHKDGSDEAHCVRLF-NGTDSGLVQFR--IQSI-----WHVACAEN 722

QY 129 FTEALAEATACRMGY-----SKPTFRAVEIGPDQDLVVE-----ITENSQELMRNWSGP 180
Db 723 WTTQISDDVCCOLLGLGTGNSVPTF-STGGGPPYVNLNAPNGSILTPSQO-----772

QY 181 CLSGSLVSLHC--LACGSLKT----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
Db 773 CLEDSLILLCNYSKCGKLVTOEVSFKIVGSGDSREGANFWVALYFDQDOVCGASLVS 832

QY 235 PHWVLTAAHC-FRKHTDFVNMKVRAGSKLGSF--PSLAVAKIIIIIEFNPVY---PKDND 288
Db 833 RDWLVSAAHCVGNWNEPSKNAVLGLHMASNLTSPTQETRLDQIVINPHYNKRKKNND 892

QY 289 IALMKLOPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLQASVOV 348
Db 893 IAMHLEMKVAVTDIYQICLPEENQVFPGRICSIAGWGALIYO-GSTADVLOEADVPL 951

QY 349 IDSTRCNADDAVQGVTEKMKCAGIEPEGVDTCQDGGSGPLMYQ--SDQHWVVGIVSGYG 407
Db 952 LSNEKCO-QQPEYNITENYVNCAGYAGVDSQCGDSGGLMCCQENRLLAGVTSFGYQ 1010

QY 408 CGGPSTPGVYTKVSAYLWNI 427
Db 1011 CALPNRPGVYARVPRFTWNI 1030

RESULT 4
S33777 hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Raymond, F.; Nick, H.
Biochem. Biophys. Acta 1173, 350-352, 1993
A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A:Reference number: S33777; PMID:93305733; PMID:8318546
A:Accession: S33777
A:Status: preliminary
A:Molecule type: mRNA
A:Cross-references: EMBL:X70900; NID:G57928; PIDN:CAA50256.1; PID:G57929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hepsin; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 568.5; DB 1; Length 416;
Best Local Similarity 30.5%; Pred. No. 1.6e-39;
Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;

QY 19 KPRIPMEFRKVGIPPIIIALLSLASIIIVWLKIVDKYFLCQPLHPIPRKQCDGE 78
Db 14 RPQVALT---VGTLLFTGIGAAWAIVTLIR-----44

QY 79 LDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALAEATAC 138
Db 45 -----SQDE-----PLYQVQLSPGDSRLILVDKTEGTWRLLCSSRNARVAGLGC 89

QY 139 ROMGY-----SSKPTFRAVEIG-----PDQDLVVEITENSQELMRNWS 178

F:801-1030/Domain: trypsin homology <TRY>
F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
F:788-912,826-942,926-993,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 23.6%; Score 552; DB 1; Length 1034;
Best Local Similarity 34.3%; Pred. No. 1.2e-37;
Matches 127; Conservative 69; Mismatches 141; Indels 34; Gaps 14;

QY 74 LDCGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEAL 133
Db 676 LDCGFSCKDGSDEAHCVR-FLNGTANNGLVQFR--IQSI-----WHTACAEINTTQT 726

QY 134 AETACROMGY-----SSKPTFRAVEIGPDQDLVVEITENSQELMRNWSGPCILSGSIVSL 189

Db 90 EENGFLRALAHSELSDVRTAGANGTSFFCVDGEGLPLAQRLLDVISCVD-----138
QY 179 GPCLSGSLVSLHCLACG-KSLKTPRVVVGGEASVDSWPQVSIQYDKQHVCGGSILDPHW 237
Db 139 --CPRGFTATQDCGRRKLPVDRIVGGQDSSLGRFPQVSLRYDGTDLGCGSLSGDW 196

QY 238 VLTAACHFRKHTDFVN-WKVRAGSDKLGSFPFSLAVAKIIIIIEFNPVY-----KDND 288
Db 197 VLTAACHCFPERNVLGRWRVFAAVARTSPHVAQLGVQAVIYHGGYLPFRDPTIDNSND 256

QY 289 IALMKLOPLTFSGTVRPICLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLQASVOV 348
Db 257 IALVHLSLPLTEYLQVCLPAAGALVDGKVCVTVGNGT--QFYGQQAVALQEARVPI 315

QY 349 IDSTRCNADDAVQGVTEKMKCAGIEPEGVDTCQDGGSGPLMYQ--SDQHWVVGIVS 403
Db 316 ISNEVCNSPDFYGNQIKPKMFCAGYPEGGIDACQDGGHFCVCEDRISGTSRWRLCGIVS 375

QY 404 WGYCGGSPSTPGVYTKVSAYLWNIYNWK 432
Db 376 WGTGSCALARKEPGVYTKVIDFREWIFQAIK 404

RESULT 5
A53663 enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N:Alternate names: enterokinase
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A53663
R:Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, I.
J. Biol. Chem. 269, 19976-19982, 1994
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; PMID:94327548; PMID:8051081
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1034 <MAT>
A:Cross-references: GB:D30799; NID:G505122; PIDN:BAA06459.1; PID:G505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from the
C:Comment: The mechanism of association with the membrane of the intestinal brush border
ated below) or with amino-terminal myristoylation of the heavy chain.
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cis repeat homology; LDL receptor ligand-binding repe
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAX homology <MAX>
F:541-646/Domain: Clr/Cis repeat homology <Clr>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF>
F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F:800-1029/Domain: trypsin homology <TRY>
F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,964
F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 23.6%; Score 552; DB 1; Length 1034;
Best Local Similarity 34.3%; Pred. No. 1.2e-37;
Matches 127; Conservative 69; Mismatches 141; Indels 34; Gaps 14;

QY 74 LDCGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEAL 133
Db 676 LDCGFSCKDGSDEAHCVR-FLNGTANNGLVQFR--IQSI-----WHTACAEINTTQT 726

QY 134 AETACROMGY-----SSKPTFRAVEIGPDQDLVVEITENSQELMRNWSGPCILSGSIVSL 189

Db 727 SDDVQLGLGTGNSMPFFSGG-GP-----FVKLNTAPNGSLILITASEQCFEDSLILL 780
 QY 190 HC--LACGKSL----KTRPVVGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAAH 243
 Db 781 QCNHKSCKGKQVAQVSPKIVGNDREGAWPVVALYNGQLLGCASLVSRSMLVSAH 840
 QY 244 C-FRKHDTVFNKVRAGSKLGF--PSLAVAKIIIIIFENPMY-----PKNDIALMKLOPP 297
 Db 841 CVYGNLEPSKWKALGLHMTSNLTPSQIVTLRDEIVINPHNRRRDXSDIAMWHLEFK 900
 QY 298 LTFSGTVRPLCPFFDEELTPATLWITGTFKQNGKMSDILLQASVQVIDSPRCNAD 357
 Db 901 VNYTDYIQICLPEENQVPPGRICSIAGKVIYQGGP-ADLOEADVPLLSNEKCO-Q 958
 QY 358 DAYQGEVTEKMKVAGIPEGVTCQDGGGLM-YQSDQHWVGVSVNGYCGGPGSTPGV 416
 Db 959 QMPENYNITENMKVAGVEGGIDSCQDGGGLMCLENNRLWLLAGVTSGYQCALPNRPGV 1018
 QY 417 YTKVSAYLNI 427
 Db 1019 YARVPKFTWI 1029

RESULT 6
 KORTPL
 Plasma kallikrein (EC 3.4.21.34) precursor - rat
 N/Alternate names: Fletcher factor; kininogenin; serum kallikrein
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
 C/Accession: A33320; S06851; I53041; S06852
 R/Beaubien, G.; Rosinski-Chupin, I.; Mattel, M.G.; Moikay, M.; Chretien, M.; Seidah, N.G.
 Biochemistry 30, 1628-1635, 1991
 A/Title: Gene structure and chromosomal localization of plasma kallikrein.
 A/Reference number: A35180; PMID:91129236; PMID:1993180
 A/Accession: A33180
 A/Molecule type: DNA
 A/Residues: 1-638 <BEA>
 A/Cross-references: GB:J05315
 A/Note: The authors translated the codon GAG for residue 81 as Gln
 R/Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur
 DNA 8, 563-574, 1989
 A/Title: The cDNA structure of rat plasma kallikrein.
 A/Reference number: A33320; PMID:90091743; PMID:2598771
 A/Accession: A33320
 A/Status: not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-638 <SEI>
 A/Cross-references: GB:M0282; NID:G205010; PIDN:AAA41463.1; PID:G205011
 A/Note: Part of this sequence, including the amino ends of both the heavy and light chain
 R/Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
 Biochim. Biophys. Acta 999, 103-110, 1989
 A/Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
 A/Reference number: S06851; PMID:90089457; PMID:2597701
 A/Accession: S06851
 A/Molecule type: protein
 A/Residues: 20-45; 391-413 <PAQ>
 R/Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazur
 DNA Cell Biol. 8, 563-574, 1989
 A/Title: The cDNA structure of rat plasma kallikrein.
 A/Reference number: I53041
 A/Accession: I53041
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-638 <RES>
 A/Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722
 C/Species: This protein, synthesized in the liver, circulates as a noncovalent complex w
 C/Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
 are linked by one or more disulfide bonds.
 C/Genetics:
 A/Genes: PK
 C/Superfamily: coagulation factor XI; trypsin homology
 C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
 F:20-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
 F:127,215,308,453,459,494/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:396/Binding site: carboxylate (Asp) (covalent) #status experimental
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.8%; Score 511; DB 1; Length 638;
 Best Local Similarity 32.0%; Pred. No. 1.7e-34;
 Matches 128; Conservative 62; Mismatches 106; Indels 104; Gaps 17;

QY 64 QPLHFIPRKOLCDG-----ELDCPL--GED--EEHCVKASF-----PEGP 98
 Db 290 EPCHF---KIYGVAFEGEELNATFVGADACQETCTKTIRCQFTYSLLPDCKAEGC 345
 QY 99 AVAVRLSKDRSTLQVLDATGNWFSACFNFTALAEATACRMGYSSKPTFRAVEIGPDQ 158
 Db 346 KCSRLRLSTGDSPTRTITYAQS-----SGYS-----371
 QY 159 DLDVTEITENSQELRMNSGPGCLSGSLVSLHCLAGKSLKTPRVVGGEEASVDSMPWQV 218
 Db 372 -LRLCKVSSD-----CTTKINA-RIVGGTNSLGEWPMQV 406
 QY 219 SIQ---YDKQHVCGGSLDPHWLTAHCFR--KHTDVFNWKKVRAG-----SKLGSFPPL 269
 Db 407 SLQWKLVSQNHMGCGSIIRQWMLTAAHCFDGPYPDV--WRIYGGILNLSEITNKTPFS 464
 QY 270 AVAKIIIIERNPMYKNDIALMKLOFLPTEGTVPEICLPFFDEBELTPATLWIIWGCF 329
 Db 465 SIKELIHQKYMSEGSYDALIKLOTPLNYTFQKPLCLPSKADNTIYTCWWTGMY 524
 QY 330 TKQGGKMSDILLQASVQVIDSTRCNADDAVQGEV-TEKMCAGIPEGVTCQDGGSGP 388
 Db 525 TKER-GETQNILKATILPVNNEC--QKKYRDYVITKQMICAGYKEGGIDACKGSDSGP 581
 QY 389 LMYQ-SDQHWVGVSVNGYCGGPGSTPGVYTKVSAYLNI 427
 Db 582 LVCKHSQWLVGITSWGEGCAKKEQPGVYTKVAEYIDWI 621

RESULT 7
 KOMSPL
 Plasma kallikrein (EC 3.4.21.34) precursor - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
 C/Accession: A36557
 R/Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont
 DNA Cell Biol. 9, 737-748, 1990
 A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compariso
 A/Reference number: A36557; PMID:9190844; PMID:2264928
 A/Accession: A36557
 A/Molecule type: mRNA
 A/Residues: 1-638 <SEI>
 A/Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
 A/Note: Part of this sequence, including the amino ends of both the heavy and light chain
 C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
 are linked by one or more disulfide bonds.
 C/Superfamily: coagulation factor XI; trypsin homology
 C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
 F:210-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
 F:391-621/Domain: trypsin homology <TRY>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.7%; Score 508; DB 1; Length 638;
Best Local Similarity 41.0%; Pred. No. 38-34; Indels 34; Gaps 10;
Matches 112; Conservative 45; Mismatches 82; Indels 34; Gaps 10;

QY 174 MRNSGFCILSGSLVSLHCLACGSLKTP-----RVVGGEEASVDSPWQVSIQ---Y 222

Db 364 MQGSSG-----YSLRLKLVDPDCTTKINARIVGGINASLGEPWQVSIQVQLV 413

QY 223 DKQHVCGGSLDPHVLTAHGER--KHTDVENKVRAG-----SKLGSFPLAVAKII 276

Db 414 SOTHLGGGSIIGRWLTAHCFDGPYPDV--WRIYGIILSLSEITKETPSRIKELII 471

QY 277 IEFNPMYKNDIALMKLPFLTFSTGTVRPICLPFFDEBLTPATPLWIIGWFTKONGK 336

Db 472 HOEYKVSSEGYDIALIKLQTLNLYTEFQKPLPSKADINTIYTCWVTGWYTKEQ-GE 530

QY 337 MSCLILQASVQVIDSTRCNADDAVOGEVTEK-MWCAGIPEGVDTCQDGSGLPLMYQ-SD 394

Db 531 TQNILOKATIPVNPNEEC--QKKYRDYVINKQMICAGYKEGTDACKGDSGLPLVCKHSG 588

QY 395 QWHVGVISWVGCGGSPSTPGVYTKVSAYLNI 427

Db 589 RNQLVGITSWGCGCKKQDPGVYTKVSEYMDWI 621

RESULT 8

JX0172
acrosin (EC 3.4.21.10) precursor form 3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
C:Accession: JX0172; JX0138
R:Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
J. Biochem. 109, 828-833, 1991
A:Title: Structure and organization of the mouse acrosin gene.
A:Reference number: JX0172; MUID:92041732; PMID:1939002
A:Accession: JX0172
A:Molecule type: DNA
A:Residues: 1-436 <WAT>
A:Cross-references: GB:S66245; NID:9238706; PID:AA820293.1; PID:9238707
R:Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
J. Biochem. 108, 785-791, 1990
A:Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its gene.
A:Reference number: JX0138; MUID:91185335; PMID:2127931
A:Accession: JX0138
A:Molecule type: mRNA
A:Residues: 4-436 <KAS>
A:Cross-references: GB:D00754; NID:9220322; PID:BA000651.1; PID:9220323
C:Comment: Acrosin is an acrosomal protease that plays an important role in the initial e-binding activity.
C:Genetics: 26/2; 95/2; 190/1; 238/3
A:Introns: 26/2; 95/2; 190/1; 238/3
A:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-42/Product: acrosin light chain #status predicted <AIC>
F:43-321/Product: acrosin heavy chain #status predicted <AHC>
F:43-286/Domain: trypsin homology <TRY>
F:22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:25-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
F:89,143,241/Active site: His, Asp, Ser #status predicted
F:235/Binding site: substrate (Asp) #status predicted

Query Match 21.6%; Score 505; DB 2; Length 436;
Best Local Similarity 40.9%; Pred. No. 3.3e-34;
Matches 106; Conservative 37; Mismatches 90; Indels 26; Gaps 7;

QY 194 CG-----KSLKTPRVVGGEEASVDSPWQVSIQY-----DKQHVCGGSLDPHVLTA 242

Db 29 CGLRFNRSQAGTRIVSGSQAQLGAWPMWVSIQIFTSNRSRYHACGGSLLNSHWLTA 88

QY 243 HCFRKHDTDFNWK-----VRAGSKDLGFPFSLA--VAKIIIEFNPMYPKNDIALMK 293

Db 89 HCFDNKKKVVYDRLVFGAQIEYGRNKFVKEPQOERYVQKIVHEKYNVTEGNDIALK 148

QY 294 LQFLTFSTGTVRPICLPFFDEBLTPATP--LWIIGWFTKQNGKMSDILLQASVQVIDS 351

Db 149 ITPVTCGNFICPCLPHF-KAGPPQIFHTCYVTGWIYKEKAPRPSFVLMEARVDLID 207

QY 352 TRCNADDAVOGEVTEKMWKAGIEGGVDTCQDGSGLPLMYQSD--OWHVGVISWVGVC 408

Db 208 DLNSTQWYNGRVTSTNVCAGYPEKIDTCQDGSGLPLMCRDNDVDFVVGITSWGVC 267

QY 409 GGPSTPGVYTKVSAYLNI 427

Db 268 ARAKRPGVYTATMDYLDWI 286

RESULT 9

T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C:Accession: T30337
R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from Xer
A:Reference number: Z20829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: EMBL:U81290; NID:92981640; PID:g2981641; PID:AA24717.1
C:Superfamily: trypsin related polyprotein; trypsin homology

Query Match 21.5%; Score 504.5; DB 2; Length 1524;
Best Local Similarity 44.0%; Pred. No. 1.7e-33;
Matches 109; Conservative 38; Mismatches 86; Indels 15; Gaps 8;

QY 194 CGKSLKT-----PRVVGEEASVDSPWQVSIQYDKOHVCGGSLDPHVLTAHCRKH 248

Db 570 CGMAPMTPKMWLPRIVGGEASPNSPWQVQIFPLRTFHCCEGAIISQWILTAAHCIRAA 629

QY 249 TDVFNKVRAGS-DKLGSPFSLAVAKIIIEFNPMYPK---DNDIALMKLQFLPTFSGTV 304

Db 630 EPSY-WTVIAGDHNRLNVESTQIRNIKTIRHDNYSYDNDIALLYLEPLDNDV 688

QY 305 RPICLPFFDEBLTPATPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVOGEV 364

Db 689 RPVCLPEPEEVLTPASVVCVTGWTGNTAEDGQPALG-LQQLPLDLSIIICNT-SYYSSEL 746

QY 365 TEKMWKAGIEG-GVDTCQDGSGLPLMYQS--DQHVGVISWVGCGGSPSTPGVYTKVS 421

Db 747 TDHMLCAGFPSSKEKDACQDGSGLPLVCQNEKQFSIYGLVSWGEGCGRVSKPGVYTKVR 806

QY 422 AYLNIYIN 429

Db 807 LFTWION 814

RESULT 10

A37344
acrosin (EC 3.4.21.10) precursor form 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
C:Accession: A37344
R:Klemm, U.; Maier, W.M.; Tsaousidou, S.; Adham, I.M.; Willison, K.; Engel, W.
Differentiation 42, 160-166, 1990
A:Title: Mouse proacrosin: cDNA sequence, primary structure and postmeiotic expression
A:Reference number: A37344; MUID:90255839; PMID:2111255
A:Accession: A37344
A:Status: preliminary
A:Molecule type: mRNA

A;Residues: 1-418 <KLE>
 A;Cross-references: GB:X52466; NID:949857; PIDN:CAA36704.1; PID:949858
 C;Superfamily: acrosin, trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase
 F;39-283/Domain: trypsin homology <TRY>
 F;18,208/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;21-151/Disulfide bonds: #status predicted
 F;25-158/Disulfide bonds: #status predicted
 F;70-86/Disulfide bonds: #status predicted
 F;85,139,238/Active site: His, Asp, Ser #status predicted
 F;175-244/Disulfide bonds: #status predicted
 F;207-223/Disulfide bonds: #status predicted
 F;234-264/Disulfide bonds: #status predicted

Query Match 21.5%; Score 503; DB 2; Length 418;
 Best Local Similarity 39.1%; Pred. No. 4.6e-34;
 Matches 108; Conservative 36; Mismatches 90; Indels 42; Gaps 7;
 QY 179 GPCSLGSLVSLHCLACGKSLKT-----GLTRQNSQACTRIVSGSAHVGAWPMWVSLQIFTSNRRY 67
 DB 23 GPC-----
 QY 226 HVGCGSILDPHWLTAHCFRKHDTDFVFNWK-----VRAGSDKLGSPF--SLAVAKIII 276
 DB 68 HACGGLLSHSHWVLTAAHCFDNKKKYDWRVLFVGAQETIYGRNKPVEPOBERYVQKIVI 127
 QY 277 IEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFDEELTPATP--LWIIGWGTQKNG 334
 DB 128 HEKYNVVTGNIDIALKLVTPVTCGNFPGCPCLPHFKAGPPRKIPHTCTVTVGWYKREA 187
 QY 335 GKMSDILQASVQVIDSTRCNADDAVQGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSD 394
 DB 188 PRPSVLMARVDLIDLCLNSTQWNGEVTSTNVACGYPECKIDTCQDSGGPLMCRDN 247
 QY 395 ---QMHVGVISNGYCGGSPSTPGVTVKVSAYLNI 427
 DB 248 ARQPFVVVGITTSWGVGCARAKRPGVYATATWYLDWI 283

RESULT 11
 S18407
 acrosin (EC 3.4.21.10) precursor - rat
 N;Contains: proacrosin
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 22-Jun-1999
 C;Accession: S18407; S30037; A56620
 R;Kleim, U.; Flake, A.; Engel, W.
 Biochim. Biophys. Acta 1090, 270-272, 1991
 A;Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic analysis
 A;Reference number: S18407; MUID:92031708; PMID:1932123
 A;Accession: S18407
 A;Molecule type: mRNA
 A;Residues: 1-437 <KLE>
 A;Cross-references: EMBL:X59254
 R;Kleim, U.; Flake, A.; Engel, W.
 submitted to the EMBL Data Library, April 1991
 A;Reference number: S30037
 A;Accession: S30037
 A;Molecule type: mRNA
 A;Residues: 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 <KLE2>
 A;Cross-references: EMBL:X59254; NID:957282; PIDN:CAA41947.1; PID:g57283
 R;Kremling, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
 DNA Seq. 2, 57-60, 1991
 A;Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
 A;Reference number: A56620; MUID:92199245; PMID:1802037
 A;Accession: A56620
 A;Status: preliminary
 A;Molecule type: DNA; mRNA
 A;Residues: 1-254, 'LCDR', 259, 'DHEL', 264, 'GRLC', 269-437 <KRE>
 A;Note: sequence modified after extraction from NCBI backbone
 A;Note: sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447, NC
 C;Superfamily: acrosin, trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen

F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-43/Domain: acrosin light chain #status predicted <LCH>
 F;43-286/Domain: trypsin homology <TRY>
 F;44-437/Domain: acrosin heavy chain #status predicted <HCH>
 F;22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;25-155,29-163/Disulfide bonds: #status predicted
 F;74-90/Disulfide bonds: #status predicted
 F;89,143,241/Active site: His, Asp, Ser #status predicted
 F;178-247/Disulfide bonds: #status predicted
 F;210-226/Disulfide bonds: #status predicted
 F;237-267/Disulfide bonds: #status predicted

Query Match 21.4%; Score 501.5; DB 2; Length 437;
 Best Local Similarity 42.7%; Pred. No. 6.5e-34;
 Matches 105; Conservative 35; Mismatches 85; Indels 21; Gaps 6;
 QY 202 RVVGGEASVDSWPMQVSIQY-----DKQVCGSILDPHWVLTAAHCFRKHDTDFVFNWK 255
 DB 42 RIVGGOTSSAWPMWVSLQIFTSNRRYHACGGLLSHSHWVLTAAHCFDNKKKYDWR 101
 QY 256 -----VRAGSDKLGSPFSLA--VAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVR 306
 DB 102 LVFGAHEIEYGRNKPVEPOBERYVQKIVHEKYNVTEGNDIALKLVTPVTCGDFVGP 161
 QY 307 ICLPFDEELTPATP--LWIIGWGTQKNGKMSDILQASVQVIDSTRCNADDAVQGEV 364
 DB 162 GCLPHF-KSGPPRIPTHCTVTVGVYKIDNAPRSPVLMARVDLIDLCLNSTQWNGRV 220
 QY 365 TEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQ---WHVGVISNGYCGGSPSTPGVTVKVS 421
 DB 221 TSTNVACGYPECKIDTCQDSGGPLMCRDTRRPFVIVGITSWGVGCARAKRPGVYATATW 280
 QY 422 AYLNWI 427
 DB 281 DYLDWI 286

RESULT 12
 S29599
 acrosin (EC 3.4.21.10) precursor - guinea pig (fragment)
 C;Species: Cavia porcellus (guinea pig)
 C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C;Accession: S29599
 R;Gerton, G.L.; Hoff, H.B.; Baba, T.
 submitted to the EMBL Data Library, May 1992
 A;Description: The amino acid sequence of guinea pig proacrosin deduced from its cDNA sequence
 A;Reference number: S29599
 A;Accession: S29599
 A;Molecule type: mRNA
 A;Residues: 1-421 <GER>
 A;Cross-references: EMBL:Z12153; NID:949559; PIDN:CAA78137.1; PID:g49560
 C;Superfamily: acrosin, trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase
 F;41-284/Domain: trypsin homology <TRY>

Query Match 21.1%; Score 493.5; DB 2; Length 421;
 Best Local Similarity 38.0%; Pred. No. 2.9e-33;
 Matches 105; Conservative 44; Mismatches 84; Indels 43; Gaps 8;
 QY 179 GPC-----LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQY-----DKQ 225
 DB 25 GPCGLRFRNLQGSV-----RIIGGQTAQPGAWPMVSLQIFMAHNNRY 69
 QY 226 HVGCGSILDPHWLTAHCFRKHDTDFVFNWKVRAGSKL---GSFPSLA-----VAKIII 276
 DB 70 HACGGILLNSHSHWVLTAAHCFDNKKKYDWRVLFVGAEEIEYGNKNKPVRAPIQERYVEKIVI 129
 QY 277 IEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFDEELTPATP--LWIIGWGTQKNG 334
 DB 130 HEKYNVTEGNDIALKLVTPVTCGNFPGCPCLPHFKAGPPRKIPHTCTVTVGWYKREA 188
 QY 335 GKMSDILQASVQVIDSTRCNADDAVQGEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQ-- 392

Db 189 PRPSVLLLEARVELIDLCLNSQWYNGVMSTNVACGYPEGKIDTCQSGGLMCRDN 248

QY 393 -SDQHHVGVISWVGCGGPGSTPGVYTKVSAVLNWI 427

Db 249 ANSPFVVGITSGVGCARAKRPGIYTATWYLDWI 284

RESULT 13

KOHUP

Plasma kallikrein (EC 3.4.21.34) precursor - human

C/Species: Homo sapiens (man)

C/Date: 13-Aug-1986 #sequence revision 13-Aug-1986 #text_change 18-Jun-1999

C/Accession: A00921; A37939

R/Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

Biochemistry 25, 2410-2417, 1986

A/Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four b

A/Reference number: A00921; MUID:86243359; PMID:3521732

A/Accession: A00921

A/Molecule type: mRNA

A/Residues: 1-638 <CHU>

A/Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263

R/McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A/Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of

A/Reference number: A37939; MUID:91152016; PMID:1998666

A/Accession: A37939

A/Molecule type: protein

A/Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-160-283,'X',285-287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',345;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCN>

C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C/Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li

C/Comment: are linked by one or more disulfide bonds.

C/Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal m

C/Comment: inogen and may also play a role in the renin-angiotensin system by converting prorenin i

C/Genetics:

A/Gene: GDB:KLK3

A/Cross-references: GDB:127575; OMIM:229000

A/Map position: 4q35-4q35

C/Superfamily: coagulation factor XI; trypsin homology

C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-638/Product: plasma kallikrein #status predicted <MAT>

F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HC>

F:20-109/Domain: apple repeat <AP1>

F:110-193/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:291-380/Domain: apple repeat <AP4>

F:391-638/Domain: plasma kallikrein light chain #status predicted <LC>

F:391-621/Domain: trypsin homology <TRY>

F:21-104;47-77;151-57;111-194;137-166;141-147;201-284;227-256;231-237;292-375;322-328;383

F:127;308;396;453;494/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:318-347;340-345/Disulfide bonds: #status predicted

F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted

F:434;483;578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 491.5; DB 1; Length 638;

Best Local Similarity 42.2%; Pred. No. 76-33; Mismatches 16; Gaps 8;

Matches 100; Conservative 47; Indels 16; Gaps 8;

QY 202 RVVGGEEASVDSWPVQVSTQYD--KQHVCGGSLDPHWLTAACHF--RKHTDVFNMKV 256

Db 390 RVGGTNSWGEMFQVSLQVLTQRLHCGSLGHQWLVTAACHFGLPLQDV--WRI 447

QY 257 RAG----SDKGSFSLAVAKIIIEFNPMPKNDIIMKLPPLTFSGTVRPICLPFF 312

Db 448 YSGIINLSDITKDTFQSKKEIIHQYKVSNGNHDIALIKIQAPLNTFEQKICLPK 507

QY 313 DEELTPATPLTWIGGFTKONGKMSDILLOASVQVIDSTRCNADDAYQG-EVTEKMWCA 371

Db 508 GTSIYINCWTWNGFSEK-GEIQNILQKNIPLVINEEC--OKRIQDIKIQRMVCA 564

QY 372 GIPEGGVDTCCQSGGGLMYQ-SDQHHVGVISWVGCGGPGSTPGVYTKVSAVLNWI 427

Db 565 GYEGGKDAKGSJSGGLVCKHGMRLVITSGGCCARREQPGVYTKVAEYMDWI 621

RESULT 14

JE0315

Low-density lipoprotein receptor-related protein - mouse

C/Species: Mus musculus (house mouse)

C/Date: 16-Jul-1999 #sequence revision 16-Jul-1999 #text_change 03-Feb-2003

C/Accession: JE0315

R/Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.

J. Biochem. 124, 784-789, 1998

A/Title: A novel low-density lipoprotein receptor-related protein with type II membrane I

A/Reference number: JE0315; MUID:98429596; PMID:9756624

A/Accession: JE0315

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1113 <TOM>

A/Cross-references: DDBJ:AB013874; NID:G3869144; PIDN:BA34371.1; PID:G3869145

C/Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligar

F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>

F:869-1097/Domain: trypsin homology <TRY>

Query Match 20.9%; Score 488.5; DB 2; Length 1113;

Best Local Similarity 33.2%; Pred. No. 2.5e-32;

Matches 130; Conservative 52; Mismatches 148; Indels 61; Gaps 17;

QY 69 IPKQLCDGELDPLGDEBHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128

Db 736 VPRDLWCDGNVDCSDSDGWGCVTLKNG-----NSSLLTVHKSKEH--HVCADG 785

QY 129 FTEALAEACRMGYSSKPTFRAVEIGPOODLDVTEITENSQELRM-----RNSG----- 179

Db 786 WRETSLQACKQMLGEPSPVTKLI---PGQ-----EGQWLRLYPNWENLNGSTLQE 834

QY 180 -----PCLSGSLVSLHCLA--CGK---SLKTRVVGGEASVDSWPVQVSIQYDKQ-HV 227

Db 835 LLVTRHSCPSRSEISLLCSKQDCGRPAARMNKRILGGRTSRGRMPWQCSLSEPSGHI 894

QY 228 CGGSILDPHWLTAACHFRKHTDVFNMKVRAAGSKL---GSFSLAVAKIIIEFNPMPY 284

Db 895 CGCVLIACKWLVTAACHFEGREDADVWKVFGINLNDHPSCFMQTRFVKTILL--HPRYS 952

QY 285 K---DNDIALMKLPPLTFSGTVRPICLPFDDELTPATPLTWIGGFTKONGKMSDIL 341

Db 953 RAVVDYDISVELSDDDINETSRYRVPCLSPSEYLEPDTYCYITGWG---HMGKMPFKL 1009

QY 342 LQASVQVIDSTRCNADDAY--QGEVTEKMWCAIGPEGVDTCQSGSGGLPMYQ--SDQWH 397

Db 1010 QEGEVRIIPLEQC---QSYFDMKTIINRMICAGYESTVDSGMSGGLVCPGQWT 1066

QY 398 VGVISWVGCGGPGST-PGVYTKVSAVLNWI 427

Db 1067 LFGLTSGVSCVCFSKVLGPGVYSNVSVFGWI 1097

RESULT 15

SL1674

acrosin (EC 3.4.21.10) precursor - human

C/Species: Homo sapiens (man)

C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000

C/Accession: SL1674; S23499; S12063; A61022; S03330

R/Keime, S.; Adham, I.M.; Engel, W.

Eur. J. Biochem. 190, 195-200, 1990

A/Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene.

A/Reference number: SL1674; MUID:90306003; PMID:2114285

A:Accession: S11674
A:Molecule type: DNA
A:Residues: 1-421 <KEI>
A:A:Cross-references: EMBL:X54017; NID:g35582; PIDN:CAA37964.1; PID:g1216165
A:A>Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue 2.
R:Vazquez-Levin, M.H.; Revientos, J.; Gordon, J.W.
Eur. J. Biochem. 207, 23-26, 1992
A:A:title: Molecular cloning, sequencing and restriction mapping of the genomic sequence e
A:Reference number: S23499; MUID:92331659; PMID:1628652
A:Accession: S23499
A:A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-421 <VAZ>
A:A:Cross-references: EMBL:M77378
A:A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
R:Keime, S.
submitted to the EMBL Data Library, December 1989
A:Reference number: S12063
A:Accession: S12063
A:Molecule type: DNA
A:Residues: 1-225, 'R', 227-421 <KSI2>
A:A:Cross-references: EMBL:X54017
R:Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A:A:title: Molecular cloning of human procacrosin cDNA.
A:Reference number: A61022; MUID:90128988; PMID:2298447
A:Accession: A61022
A:A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A:A:title: Primary structure of human procacrosin deduced from its cDNA sequence.
A:Reference number: S03330; MUID:89153568; PMID:2493394
A:Accession: S03330
A:Molecule type: mRNA
A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
A:A:Cross-references: EMBL:Y00970; NID:g28325; PIDN:CAA68784.1; PID:g28326
C:Genetics:
A:Gene: GDB:ACR
A:A:Cross-references: GDB:119645; OMIM:102480
A:Map position: 22q13-22qter
A:Introns: 26/2; 94/2; 189/2; 237/3
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-19/DNA: signal sequence #status predicted <SIG>
F:20-421/Product: acrosin #status predicted <MAT>
F:20-42/Product: acrosin light chain #status predicted <LCH>
F:43-421/Product: acrosin heavy chain #status predicted <HC>
F:43-285/DNA: trypsin homology <TRY>
F:302-379/Region: proline-rich
F:22,210/Binding site: carbonyl hydrate (Asn) (covalent) #status predicted
F:25-154/Disulfide bonds: #status predicted
F:29-162/Disulfide bonds: #status predicted
F:73-89/Disulfide bonds: #status predicted
F:88,142,240/Active site: His, Asp, Ser #status predicted
F:177-246/Disulfide bonds: #status predicted
F:209-225/Disulfide bonds: #status predicted
F:236-266/Disulfide bonds: #status predicted

162	CLPHFKAGLPGSSQSCVWAGWGVIIEKAPRPSIIEMEARVDLIDLDCNSTOWYNGRVP	221
dbb		
367	KMCAGIPEGGVDTCCGDSGGPLNYGSDQ---WHVVGIVSWVGCGCGPSTPGVYTKVAS	423
QY		
222	TNVCAGIPVGIKIDITCCGDSGGPLMCKDSKESAYVVVGI-TSWVGICALAKRPGIVATWPE	281
dbb		
424	LNWI	427
QY		
282	LNWI	285
dbb		

Search completed: June 7, 2004, 08:41:05
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 18 Seconds
(without alignments)
1258.362 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVPLKRP.....VYTKVSAYLNWYNWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2337	99.8	437	1 TMS4 HUMAN	Q9nr54 homo sapien
2	1844	78.7	435	1 TMS4 MOUSE	Q8vca5 mus musculus
3	687	28.3	453	1 TMS3 MOUSE	Q8k1t0 mus musculus
4	673.5	28.8	492	1 TMS2 HUMAN	Q15393 homo sapien
5	660.5	28.2	454	1 TMS3 HUMAN	P57727 homo sapien
6	648.5	27.7	490	1 TMS2 MOUSE	Q9j1q8 mus musculus
7	628.5	26.8	455	1 TMS5 MOUSE	Q9er04 mus musculus
8	584.5	25.0	417	1 HEP5 HUMAN	P05981 homo sapien
9	582.5	24.9	1019	1 ENTK HUMAN	P98073 homo sapien
10	578	24.7	436	1 HEP5 MOUSE	Q35453 mus musculus
11	574	24.5	457	1 TMS5 HUMAN	Q9h3s3 homo sapien
12	571	24.4	1035	1 ENTK BOVIN	P98072 bos taurus
13	568.5	24.3	416	1 HEP5 RAT	Q05511 rattus norv
14	568	24.3	418	1 HATT HUMAN	O60235 homo sapien
15	552.5	23.6	1069	1 ENTK MOUSE	P97435 mus musculus
16	552	23.6	1034	1 ENTK PIG	P98074 sus scrofa
17	550.5	23.5	811	1 TMS6 MOUSE	Q9dbi0 mus musculus
18	531	22.7	811	1 TMS6 HUMAN	Q8tu80 homo sapien
19	511	21.8	638	1 KAL RAT	P14272 rattus norv
20	508	21.7	638	1 KAL MOUSE	P26262 mus musculus
21	505	21.6	436	1 ACRO MOUSE	P23578 mus musculus
22	503.5	21.5	324	1 TEST MOUSE	Q91bj7 mus musculus
23	501.5	21.4	437	1 ACRO RAT	P22933 rattus norv
24	498	21.3	875	1 NETR HUMAN	P03952 homo sapien
25	491.5	21.0	638	1 KAL HUMAN	Q93219 mus musculus
26	488.5	20.9	1113	1 CORI HUMAN	Q9y5q5 homo sapien
27	484	20.7	1042	1 CORI HUMAN	Q9y5q5 homo sapien
28	483	20.6	321	1 TRYG HUMAN	Q9nr12 homo sapien
29	482	20.6	855	1 ST14 MOUSE	P56677 mus musculus
30	481	20.5	421	1 ACRO HUMAN	P10323 homo sapien
31	478.5	20.4	415	1 ACRO PIG	P08001 sus scrofa
32	478	20.4	314	1 TEST HUMAN	Q9y6m0 homo sapien
33	476.5	20.3	311	1 TRYG MOUSE	Q9qul7 mus musculus

34	475	20.3	422	1 DES1 HUMAN	Q9ul52 homo sapien
35	472.5	20.2	317	1 BSS4 HUMAN	Q9gn44 homo sapien
36	470	20.1	751	1 NETR MOUSE	Q08762 mus musculus
37	468.5	20.0	431	1 ACRO RABIT	P48038 oryctolagus
38	467.5	20.0	290	1 PR27 HUMAN	Q9bgr3 homo sapien
39	463	19.8	855	1 ST14 HUMAN	Q9y5v6 homo sapien
40	461.5	19.7	277	1 TRY2 ANOGA	P35036 anopheles g
41	460.5	19.7	263	1 CTRB RAT	P07338 rattus norv
42	459.5	19.6	342	1 PSS8 RAT	Q9e887 rattus norv
43	459.5	19.6	812	1 PLMN MOUSE	P20918 mus musculus
44	458.5	19.6	267	1 TRY7 ANOGA	P35041 anopheles g
45	458.5	19.6	790	1 PLMN PIG	P06867 sus scrofa

ALIGNMENTS

RESULT 1
TMS4 HUMAN
ID TMS4 HUMAN STANDARD; PRT; 437 AA.
AC Q9NR54; Q9NZAS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
protease 2) (MT-SP2).
GN TMPSR94 OR TMPSR53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMPSR53) overexpressed in
pancreatic cancer."
RL Cancer Res. 60:2602-2606(2000).
[2]
RN RN
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
ENAc (By similarity).

CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: High levels in pancreatic, gastric, colorectal
CC and ampullary cancer. Very weak expression in normal
CC gastrointestinal and urogenital tract.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF179224; AAF74526.1; -;
CC EMBL; AF216312; AAF31436.1; -;
CC EMBL; BC011703; AAF11703.1; -;
CC HSP; P00763; IDPO.
CC MEROPS; S01.034; -;
CC Genew; HGNC:11678; TMPRSS4.
CC MIM; 605563; -;
CC GO; GO:0016021; C:integral to membrane; NAS.
CC GO; GO:0004252; P:serine-type endopeptidase activity; NAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR002172; LDL receptor A.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001314; Peptidase_S1a.
CC Pfam; PF00057; ldl_recept_a; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00192; LDLa; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00202; Tryp_Spc; 1.
CC PROSITE; PS01209; LDLRA_1; FALSE NEG.
CC PROSITE; PS00668; LDLRA_2; FALSE NEG.
CC PROSITE; PS00420; SRCR_1; FALSE NEG.
CC PROSITE; PS0287; SRCR_2; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 33 53 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 33 53 (POTENTIAL).
FT DOMAIN 54 437 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 61 93 LDL-RECEPTOR CLASS A.
FT DOMAIN 94 204 SRCR.
FT DOMAIN 205 437 SERINE PROTEASE.
FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 204 205 CLEAVAGE (POTENTIAL).
FT DISULFID 64 83 BY SIMILARITY.
FT DISULFID 77 92 BY SIMILARITY.
FT DISULFID 127 183 BY SIMILARITY.
FT DISULFID 140 193 BY SIMILARITY.
FT DISULFID 196 310 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 356 372 BY SIMILARITY.
FT DISULFID 383 410 BY SIMILARITY.
FT CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 1 31 NQDPSDQPLNSLDVPLKRPDMETFRK -> MSNPCA
FT CONFLICT 1 31 MFQDPSDPSSES (IN REF. 2).
FT SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
Query Match 99.8%; Score 2337; DB 1; Length 437;
Best Local Similarity 100.0%; Pred. No. 4.2e-194; Indels 0; Gaps 0;
Matches 434; Conservative 0; Mismatches 0;

QY 2 DPDSQPLNSLDVPLKRPDMETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVPLKRPDMETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGOPLHFIPRKQCDGELDCPLGEDEEHCHVKSPEGPAVAVRLSKDRSTLQVLD SATGNW 121
DB 64 CGOPLHFIPRKQCDGELDCPLGEDEEHCHVKSPEGPAVAVRLSKDRSTLQVLD SATGNW 123
QY 122 FSACFDFNTEALAEACRQMGYSKTFRAVEIGPDODLDVVEITENSQELMRNSSGPC 181
DB 124 FSACFDFNTEALAEACRQMGYSKTFRAVEIGPDODLDVVEITENSQELMRNSSGPC 183
QY 182 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHVLT 241
DB 184 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHVLT 243
QY 242 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFPMYKNDIIMKLOFLTFS 301
DB 244 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFPMYKNDIIMKLOFLTFS 303
QY 302 GTVRPICLPFFDEELTPATPLWIGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 361
DB 304 GTVRPICLPFFDEELTPATPLWIGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 363
QY 362 GEVTEKMKCAGIPGGVDTTCQSDSGPLMYQSDQHWVGVISVWGVCSPGVTYKVS 421
DB 364 GEVTEKMKCAGIPGGVDTTCQSDSGPLMYQSDQHWVGVISVWGVCSPGVTYKVS 423
QY 422 AYLNWIYVWKAEL 435
DB 424 AYLNWIYVWKAEL 437
RESULT 2
TMS4 MOUSE
ID TMS4 MOUSE STANDARD; PRT; 435 AA.
AC Q8VCA5; 2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Channel-activating
DE protease 2) (mCAP2).
GN TMPRSS4 OR CAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22144321; PubMed=12149280;
RA Vuagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
RT "Synergistic activation of ENac by three membrane-bound channel-
RT activating serine proteases (mCAP1, mCAP2, and mCAP3) and serum-
RT glucocorticoid-regulated kinase (Sgk1) in Xenopus oocytes.";
RL J. Gen. Physiol. 120:191-201(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Breast tumor;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

[illegible]

```
DR EMBL; AJ429216; CAD22137.1; -
DR EMBL; AJ300738; CAC83350.1; -
DR HSP; P00761; IAN1.
DR MGD; MG1:2155445; Tmpres3.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_receptor_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00688; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin SR; 1.
DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Endoplasmic reticulum.
FT DOMAIN 1 48
FT TRANSMEM 49 69
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT POTENTIAL.
FT FT 70 453
FT DOMAIN 72 108
FT DOMAIN 104 205
FT DOMAIN 217 448
FT ACT_SITE 257 257
FT ACT_SITE 304 304
FT ACT_SITE 400 400
FT SITE 216 217
FT DISULFID 73 85
FT DISULFID 79 98
FT DISULFID 92 107
FT DISULFID 129 184
FT DISULFID 142 204
FT DISULFID 207 324
FT DISULFID 242 258
FT DISULFID 338 406
FT DISULFID 369 385
FT DISULFID 396 424
FT CARBOHYD 221 221
FT CONFLICT 117 117
FT CONFLICT 246 246
FT SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1EF6 CRC64;
Query Match 29.3%; Score 687; DB 1; Length 453;
Best Local Similarity 36.9%; Pred. No. 1.2e-51;
Matches 161; Conservative 72; Mismatches 153; Indels 50; Gaps 14;
QY 22 IPMETFRKVGIPITIIALLSLASIIIVVLKIVLD---KYFLCQQLHFIPRKQLCDGE 78
DB 44 LPLKFFPIIVIGIALLALA-----ICLGHPDCGKY--RCHSSFKIELTARCDGV 95
QY 79 LDCPLGEDEHCKVFPPEGPAVAVPLSDKDRSTLQVLDATGNWSACPDNTEALAEAC 138
DB 96 SDCKNADEYRC-----VRVSGRAALQVFTAAA--WRTMCSDDDKSHYAKIAC 142
QY 139 ROMGYSS-----KPTFRAVEICPDQDLWDWEITENSQELRNRSSGPCLSGL 186
DB 143 AQLGFPYSVSSDLHRLVDALEEFQGFDFVSIHLLSDKRVTAHLSVYMRG--CTSGHV 199
QY 187 VSLHCLAG-KSLKTPRVVGGEEASVDGAPQVSTQYDKQVCGGSLDPHWVTAHCF 245
DB 200 VTLKCSAGTRTGYSPRIVGNNMSLTQMPQVSLQFGQYHLCGSGVITPLWIVTAHCV 259
QY 246 RKHTDVFNNKVRAGSKL--GSFPLAVAKIITIEFNPMY-PK--DNDIALMKLQFPITF 300
DB 260 YDLHYPKSWTVQVGLVSLMDSVPVPSHLEKII---YHSKYFKRLGNDIALMKLSEPLTF 316
```

```
QY 301 SGTVRPICLPFFDEELTPATPLWIGWFTKQNGKMSDILLOASVOVIDSTRCNADDAY 360
DB 317 DETIOPICLPNSENFPDGKLCWTSGWAT-EGGDASPVNLHAAVPLISNKICNHRDVI 375
QY 361 QGEVTERKMCAG:PEGGVDTCQDSSGGLMYQSDQ-WHVVGVISWVGCGGPGSTPGVYTK 419
DB 376 GGIISPSMLCAGYLKGGVDSQDSSGGLVCQERRLWLKLVGATSGFICCAEVNKPQVYTR 435
QY 420 VSAVLNWIYVWKAEL 435
DB 436 ITSFLDWIHEQLERDL 451
RESULT 4
TMS2_HUMAN
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC OLS393; Q9BX11;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=97458144; PubMed=9325052;
PAoloni-giacobino A., Chen H., Feitsch M.C., Rossier C.,
Antonarakis S.E.;
"Cloning of the TMRSS2 gene, which encodes a novel serine protease
with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";
Genomics 44:309-320(1997).
[2]
SEQUENCE FROM N.A.
MEDLINE=21309069; PubMed=11414763;
Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
"Mutation analyses of 268 candidate genes in human tumor cell lines.";
Genomics 74:352-364(2001).
[3]
SEQUENCE FROM N.A. AND MUTAGENESIS.
MEDLINE=21139112; PubMed=11245484;
Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,
Raitano A.B., Jakobovits A.;
"Catalytic cleavage of the androgen-regulated TMRSS2 protease results
in its secretion by prostate and prostate cancer epithelia.";
Cancer Res. 61:1686-1692(2001).
[4]
TISSUE SPECIFICITY
MEDLINE=21104370; PubMed=11169526;
Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikho P.T.;
"Expression of transmembrane serine protease TMRSS2 in mouse and
human tissues.";
J. Pathol. 193:134-140(2001).
-1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
cleavage and secreted.
-1- TISSUE SPECIFICITY: Expressed strongly in small intestine. Also
expressed in prostate, colon, stomach and salivary gland.
-1- SIMILARITY: Belongs to peptidase family S1.
-1- SIMILARITY: Contains 1 LDL-receptor class A domain.
-1- SIMILARITY: Contains 1 SRCR domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
EMBL; U75329; AAC51784.1; -
EMBL; AF123453; AAD31117.1; -
```


DR EMBL; AF270487; AAK29280.1; --
 DR HSP60; P00763; IDPO.
 DR MEROPS; S01.247; --
 DR Genew; HGNC:11876; TMPRSS2.
 DR MIM; 602060; --
 DR GO; GO:0003887; C: integral to plasma membrane; TAS.
 DR GO; GO:0008236; F: serine-type peptidase activity; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00668; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS00287; SRCR_2; 1.
 DR PROSITE; PS00240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Signal-anchor; Zymogen;
 KW Polymorphism.
 FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
 FT CHAIN 256 492 CATALYTIC CHAIN.
 FT CHAIN 492 TRANSMEMBRANE PROTEASE, SERINE 2,
 FT CHAIN 492 CATALYTIC CHAIN.
 FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 112 149 LDL-RECEPTOR CLASS A.
 FT DOMAIN 150 242 SRCR.
 FT DOMAIN 256 492 SERINE PROTEASE.
 FT ACT_SITE 296 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 441 441 CHARGE RELAY SYSTEM.
 FT SITE 255 256 CLEAVAGE (POTENTIAL).
 FT DISULFID 113 126 BY SIMILARITY.
 FT DISULFID 120 139 BY SIMILARITY.
 FT DISULFID 133 148 BY SIMILARITY.
 FT DISULFID 172 231 BY SIMILARITY.
 FT DISULFID 185 241 BY SIMILARITY.
 FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 281 297 BY SIMILARITY.
 FT DISULFID 410 426 BY SIMILARITY.
 FT DISULFID 437 465 BY SIMILARITY.
 FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 449 449 K -> N (in dbSNP:1056602).
 FT /FTId=VAR_011692.
 FT MUTAGEN 255 R->Q: LOSS OF CLEAVAGE.
 FT MUTAGEN 441 S->A: LOSS OF ACTIVITY.
 FT CONFLICT 160 160 M -> V (in REF. 3).
 FT CONFLICT 242 242 I -> L (in REF. 1).
 FT CONFLICT 329 329 E -> Q (in REF. 1).
 FT CONFLICT 489 491 RAD -> KAN (in REF. 1).
 FT SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
 Query Match 28.8%; Score 673.5; DB 1; Length 492;
 Best Local Similarity 38.8%; Pred. No. 28-50;
 Matches 149; Conservative 58; Mismatches 128; Indels 49; Gaps 13;
 75 CDGELDCPLGEDEHCVCSPFPGPAVAVLSKDRSTLQVLDSATGNWFSACFDNFTEALA 134
 133 CDGVSHCPGGEDEHRCVRLY-GP-----NFIQWSSQKSNHPCVODDWNENYG 181
 135 ETACRQMGVSSKPTTRAVIEGPDQLDVLVEITENSCELMRNSSG-----PCL 182
 182 RAACRDMGV--KNNFYSSQ-----GIVDDSGSTSPFKLNTSAGNVDVYKLYHSDACS 232

QY 183 SGLVSLVHLACGKSL---KTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWVL 239
 DB 233 SKAVVSLRCIACGVNLNLSRQSRIVGGSALPGAWPQVSLHVQNVHVCVGSIIITPEWV 292
 QY 240 TAAHCFKPKH-TDVFNVKVRAGSKDLGSF---PSLAVAKIIITFENPMY---PKNDIAL 291
 DB 293 TAAHCFKPKH-TDVFNVKVRAGSKDLGSF---PSLAVAKIIITFENPMY---PKNDIAL 348
 QY 292 MKLQFPLTSGTVPICLPFFDEELTPATLWIGNGFTKQNGKMSDILLQASVQVIDS 351
 DB 349 MKLQKPLTFLNDLVKPVCLPNPQWMLQPEQLCWISGNGATEEK-GKTSEVLNAKVLIIET 407
 QY 352 TRCNADDAAYQGEVTEKMKACGIPGGVDTGCGDSGGPLMYQSDQ-WHVVGIVSGYGGG 410
 DB 408 QRCNSRYVYDNLITPAMICAGFLQGNVDSCQSGSGPLVTSKNNIWLIGDTSGWGCAC 467
 QY 411 PSTEGVYTKVSYALNWIYVWKA 434
 DB 468 AYREGVYGVNVFTDVIYRQWRAD 491
 RESULT 5
 TMS3 HUMAN
 ID TMS3 HUMAN STANDARD; PRT; 454 AA.
 AC P57727;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 TAGD-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPRSS3 OR TAGD12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian Carcinoma;
 RX MEDLINE=20521358; PubMed=11069177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Parmley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 protease."
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 [2]
 SEQUENCE FROM N.A. (ISOFORMS A; B AND D), AND VARIANT ILE-53.
 RX MEDLINE=20578749; PubMed=1137999;
 RA Scott H.S., Kudon J., Wattenhofer M., Shibuya K., Berry A., Charat R.,
 RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Younis F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C.,
 RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
 RA Antonarakis S.E.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 protease causing both congenital and childhood onset autosomal
 recessive deafness."
 RL Nat. Genet. 27:59-63(2001).
 [3]
 SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
 RX MEDLINE=22281255; PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Doughty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
 RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudon J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
 RT "The transmembrane serine protease (TMPRSS3) mutated in deafness
 DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro."
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 [4]
 RP VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
 RX MEDLINE=21354462; PubMed=11462234;
 RA Masmoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,
 RA Papasavvas M.P., Drira M., Elgaid-Bouilla A., Wattenhofer M.,
 RA Rossier C., Scott H.S., Ayadi H., Guipponi M.,

RT "Novel missense mutations of TMPPRS3 in two consanguineous Tunisian families with non-syndromic autosomal recessive deafness.";
 RL Hum. Mutat. 18:101-108(2001).
 RN [5]
 RP VARIANTS DFNB8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS
 RP ILE-53; SER-111 AND VAL-253.
 RX MEDLINE-21317610; PubMed-11424922;
 RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
 RA Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,
 RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
 RA Hance W.E., Wilcox E.R., Friedman T.B., Morell R.J.
 RT "Novel mutations of TMPPRS3 in four DFNB8/B10 families segregating
 RT congenital autosomal recessive deafness.";
 RL J. Med. Genet. 38:396-400(2001).
 RN [6]
 RP VARIANTS DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
 RX MEDLINE-21904597; PubMed-11907649;
 RA Wattenhofer M., Di Iorio V., Rabionet R., Dougherty L., Pampanos A.,
 RA Schwede T., Montserrat-Sentis B., Arbones L., Iliades T.,
 RA Pasquadi-Bisceglie A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
 RA Petersen M.B., Estivill X., Gasparini P., Scott H.S.,
 RA Antonarakis S.E.;
 RT "Mutations in the TMPPRS3 gene are a rare cause of childhood
 RT nonsyndromic deafness in Caucasian patients.";
 RL J. Mol. Med. 80:124-131(2002).
 CC -1- FUNCTION: Probable protease. Seems to be capable of activating
 CC ENAC.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Name=A;
 CC IsoId=P57727-1; Sequence=VSP_005391;
 CC Name=B; Synonyms=C;
 CC IsoId=P57727-2; Sequence=VSP_005391;
 CC Name=D;
 CC IsoId=P57727-3; Sequence=VSP_005392;
 CC Name=T; Synonyms=Truncated TAG-12V;
 CC IsoId=P57727-4; Sequence=VSP_005393; VSP_005394;
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues including fetal
 CC cochlea. Isoform T is found at increased levels in some
 CC carcinomas.
 CC -1- PTM: Undergoes autoproteolytic activation.
 CC -1- DISEASE: Defects in TMPPRS3 are a cause of childhood-onset
 CC autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
 CC -1- DISEASE: Defects in TMPPRS3 are a cause of congenital autosomal
 CC recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AR20380; AG37012.1; -;
 CC EMBL; AB038157; BAB20077.1; -;
 CC EMBL; AB038158; BAB20078.1; -;
 CC EMBL; AB038159; BAB20079.1; -;
 CC EMBL; AB038160; BAB20080.1; -;
 CC HSSP; P00763; 1DPO.
 CC XEROPS; S01.079; -;
 CC Genew; HGNC:11877; TMPPRS3.
 CC MIM; 605511; -;
 CC MIM; 601072; -;
 CC MIM; 605316; -;
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
 CC GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; SRCR_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW Endoplasmic reticulum; Deafness; Alternative splicing;
 KW Disease mutation; Polymorphism.
 FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 49 69 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 70 454 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 72 108 LDL-RECEPTOR CLASS A.
 FT DOMAIN 109 205 SRCR.
 FT DOMAIN 217 454 SERINE PROTEASE.
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 401 401 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 216 217 CLEAVAGE (POTENTIAL).
 FT DISULFID 73 85 BY SIMILARITY.
 FT DISULFID 79 98 BY SIMILARITY.
 FT DISULFID 92 107 BY SIMILARITY.
 FT DISULFID 129 194 BY SIMILARITY.
 FT DISULFID 142 204 BY SIMILARITY.
 FT DISULFID 207 324 BY SIMILARITY.
 FT DISULFID 242 258 BY SIMILARITY.
 FT DISULFID 338 407 BY SIMILARITY.
 FT DISULFID 370 386 BY SIMILARITY.
 FT DISULFID 397 425 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 1 127 Missing (in isoform B).
 FT VARSPLIC 318 454 Missing (in isoform T).
 FT VARSPLIC 261 293 Missing (in isoform T).
 FT VARSPLIC 294 454 Missing (in isoform T).
 FT VARIANT 53 53 V -> I.
 FT VARIANT 103 103 D -> G (in DFNB8/DFNB10).
 FT VARIANT 109 109 R -> W (in DFNB8/DFNB10).
 FT VARIANT 111 111 G -> S.
 FT VARIANT 173 173 D -> N.
 FT VARIANT 194 194 C -> F (in DFNB8/DFNB10).
 FT VARIANT 251 251 W -> C (in DFNB8/DFNB10).
 FT VARIANT 253 253 I -> V (in absNP:2839500).
 FT VARIANT 253 253 /FTId=VAR_013101.

```

FT VARIANT 404 404 P -> L (in DFN8/DFN10).
FT VARIANT 407 407 /FTID=VAR_011679.
FT VARIANT 407 407 C -> R (in DFN8/DFN10).
FT VARIANT 426 426 /FTID=VAR_013495.
FT VARIANT 426 426 A -> T.

Query Match 28.2%; Score 660.5; DB 1; Length 454;
Best Local Similarity 35.4%; Pred. NO. 2.4e-49;
Matches 157; Conservative 74; Mismatches 149; Indels 63; Gaps 17;

QY 22 IPMEFRKVGPIPIIALSLASIIIVVLKIVLD---KY---YELGQGLHPIPKQL 74
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
44 LPLKEFPPIVIGIILIIILALA-----LGLIHDFCSGKYCRSFKC---IELIAR--- 91
QY 75 CDGELDCPLGDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATCNWFSACDFNFTALA 134
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92 CDGVSDCKDGEDEYRC-----VRVGQNAVILQVFTAA--SWKTMCSDDWKGHYA 138
QY 135 ETACHOMGYSS-----KPTFRVAVIGPDLDVVEITENSQELMRNNSGCL 182
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 NVACAQLGFPYSVSDNIRVSLLEGQFREBFVSDHLLPDDKVTALHHSVTVREG---CA 195
QY 183 SGLSVSLHCLAG-KSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWILTA 241
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 SGHVTILOCTACCHRGYSSRIVGNMSLLSQWPAQLQGGYHLGSGVITPLMIITA 255
QY 242 AHCFRKHDTVF---NWKVRAGSKL--GSFPLSLAVAKIIIIIEFNPMY-PK--DNDIALMK 293
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 ARCV---YDLPLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPELGNLIALMK 309
QY 294 LQFPLTFSTVRPICLPFFDEBELTATPLWIIGWFTKQNGKWSLILQASVOVIDSTR 353
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
310 LAGPLTFNEMQVCLPNSSENFPGKVCWTSGWAGTGDGADGASVPLNHAAPLISNKI 369
QY 354 CNADDAYGQVETKMKACIGPEGGYDTQCGSGGFLAMYQSDQ-MHVVGIVSGWGGCGPS 412
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
370 CNHRDVGGLIISPSMLCAGYLTTGGVDSQCGSGGFLVQCERRLKLKLVGATSGFGICAEVN 429
QY 413 TPGVTVKVSAYLNWYINWKAEL 435
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
430 KPGVTVRTVSLFDLWHEQMERDL 452

RESULT 6
TMS2 MOUSE STANDARD; PRT; 490 AA.
ID AC TMS2_MOUSE Q9UJK4; Q9Y82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE Transmembrane protein X).
GN TMRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c;
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;

```

```

RX MEDLINE=20148617; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin.";
RL FEBS Lett. 468:93-100 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan A., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
CC cleavage and secreted (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed mainly in prostate and kidney.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF199362; AAF97867.1; -
CC EMBL; AF243500; AAP64186.1; -
CC EMBL; AF113596; AAP21308.1; -
CC EMBL; BC038393; AAH38393.1; -
CC HSSP; P00763; IDPO.
CC MEROPS; S01.247; -.
CC MGD; MGI:1354381; Tmprss2.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00057; ldl_recept_a; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00083; tryptsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00192; LDLa; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00202; Tryp_Spc; 1.
CC PROSITE; PS01209; LDLRA_1; 1.
CC PROSITE; PS00068; LDLRA_2; 1.
CC PROSITE; PS00420; SRCR_1; FALSE_NEG.
CC PROSITE; PS00287; SRCR_2; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS0287; TRYPSIN_HIS; 1.
CC PROSITE; PS00134; TRYPSIN_SER; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.

```

```
FT CHAIN 254 490 TRANSMEMBRANE PROTEASE, SERINE 2,
FT FT CATALYTIC CHAIN.
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT TRANSMEM 84 104 (POTENTIAL).
FT FT
FT DOMAIN 105 490 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 111 149 LDL-RECEPTOR CLASS A.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 254 490 SERINE PROTEASE.
FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 253 254 CLEAVAGE (POTENTIAL).
FT DISULFID 76 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 171 230 BY SIMILARITY.
FT DISULFID 184 240 BY SIMILARITY.
FT DISULFID 243 363 INTERCHAIN (BY SIMILARITY).
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 424 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 S -> L (IN REF. 3).
FT CONFLICT 122 122 S -> N (IN REF. 3).
FT CONFLICT 178 178 Y -> H (IN REF. 1).
FT CONFLICT 320 320 Y -> H (IN REF. 1).
FT CONFLICT 474 474 N -> D (IN REF. 1).
FT SEQUENCE 490 AA; 53479 MW; 07D2B03EAD8A19 CRC64;

Query Match 27.7%; Score 648.5; DB 1; Length 490;
Best Local Similarity 35.2%; Pred. No. 2,9e-48;
Matches 161; Conservative 63; Mismatches 175; Indels 59; Gaps 15;

QY 16 PLRKPRIPIETKRV-----GIPI-----IALSLASIIIVVLTKVILDKY- 59
DB 50 PQYAPRIITQASTVTHTPKSSGAPCTSKSKSLCLALGTVLTAAGAVALLLREWD 109
QY 60 -----FLCGQPLHPIPRKQCLDGLDCEDEEHCVKSPPEGFAVAVRLSKDRSTLQ 112
DB 110 SNCSTSEMECGSGTICISSSLWCDGVANCPNGEDENRC-----VRLYGQSFIQ 158
QY 113 VLDATGNWFSAFNFTALAEATACROMYSGKPTFAVEIGPQD--LDVVEITENSQ 170
DB 159 VYSSORKAWPYCODMSESYGRAACKDMGY--KNFYSSQGIPOQSGATSFMKLNVSQ 216
QY 171 ELRMENS---SGPCLSLSVLSLHCLACG-KSLK-TPRVVGGEEASVDSWPMQVSTQYDKQ 225
DB 217 NVDLKYLHSDSCSRMVVSLRCEGVRSVRSRIUGNLNAPGDPWPQVSLHVGQV 276
QY 226 HVCGGSILDPHWLTAACFR-----KHTDVFNWVKVAGSDKLGSPFLSAVAKIIIEF 279
DB 277 HVCGGSIIIPETIVTAACHVEEPLSGPRYWTAFAGILRQSLMFYGS--RHQVEKVI---S 331
QY 280 NPMY---PKNDIAMKLOFLTFESGTVRPICLPFDEELPATPLWILIGGFTKQNGK 336
DB 332 HENYDSKTKNDIAMKLGQTPFLAFNDLVKPVCLPNPQWMLDQECVIGSWGATVEK-GK 390
QY 337 MDGILLQASQVVIDTRCNADDAVOGEVTEKMKACAGIEGGGVDTQCGSDGGPLM-YQSDQ 395
DB 391 TSDVLNAAWVPLIEFSKNSKYIYNLITPAMICAGFLQSGVDSQCGSGGGLVTLKNGI 450
QY 396 WHVVGIVSGVCGGSPSPGVTKVSAYNLVYNWKA 433
DB 451 WVLIGDTSWGCACALREGVYGNVTFVTDWYIQMRA 488
```

RESULT 7

TMS5_MOUSE

ID TMS5_MOUSE STANDARD; PRT; 455 AA.

AC Q9ER04; Q9ER02; Q9ER03;

16-OCT-2001 (Rel. 40, Created)

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
OS TMPSRSS.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "cDNA cloning of mouse spinesin.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of mouse type 4 spinesin.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q9ER04-1; Sequence=VSP_005397; VSP_005398;
CC Name=2;
CC IsoId=Q9ER04-2; Sequence=VSP_005397; VSP_005398;
CC Name=3;
CC IsoId=Q9ER04-3; Sequence=VSP_005395;
CC Name=4;
CC IsoId=Q9ER04-4; Sequence=VSP_005396;
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB016229; BAB20276.1; -
CC EMBL; AB016230; BAB20277.1; -
CC EMBL; AB016423; BAB20278.1; -
CC EMBL; AB041037; BAB40328.1; -
CC HSP; P00763; IDPO.
CC MEROPS; S01.313; -.
CC MGD; MGI.1933407; TmpRSS5.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYD_SPC; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC PROSITE; PS00420; SRCR_1; FALSE_NEG.
CC PROSITE; PS00287; SRCR_2; 1.
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor;
CC Glycoprotein; Alternative splicing.
CC DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC
CC DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 112 207 SRCR.
CC DOMAIN 218 455 SERINE PROTEASE.
CC ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SITE 217 218 CLEAVAGE (POTENTIAL).
```

"A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells.";
 Biochemistry 27:1067-1074(1988).
 [2]
 SEQUENCE FROM N.A.
 RP TISSUE=Pancreas, and Spleen;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan A., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McWay P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smillus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RN CHARACTERIZATION.
 RP MEDLINE=91358502; PubMed=1895621;
 RX Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
 RA Chou S.H., Kurachi K.;
 RA "Hepsin, a cell membrane-associated protease. Characterization,
 RT tissue distribution, and gene localization.";
 RT J. Biol. Chem. 266:16948-16953(1991).
 RN [4]
 RN CHARACTERIZATION.
 RP MEDLINE=93348237; PubMed=8346233;
 RX Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
 RA "Hepsin, a putative cell-surface serine protease, is required for
 RT mammalian cell growth.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
 CC -1- FUNCTION: Plays an essential role in cell growth and maintenance
 CC of cell morphology.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: Present in most tissues, with the highest
 CC level in liver.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M18930; AAA36013.1; -;
 DR EMBL; X07732; CAA30058.1; -;
 DR EMBL; X07002; CAA30058.1; -;
 DR EMBL; BC025716; AAH25716.1; -;
 DR PIR; S00845; S00845.
 DR HGSP; P00763; IDPO.
 DR MEROPS; S01.224; -;
 DR Genew; HGNC:5155; HPN.
 DR MIM; 142440; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.


```

CC      EMBL; U09860; AAC50138.1; -
DR      EMBL; Y19124; CAB65555.1; -
DR      EMBL; Y19125; CAB65555.1; JOINED.
DR      EMBL; Y19126; CAB65555.1; JOINED.
DR      EMBL; Y19127; CAB65555.1; JOINED.
DR      EMBL; Y19128; CAB65555.1; JOINED.
DR      EMBL; Y19129; CAB65555.1; JOINED.
DR      EMBL; Y19130; CAB65555.1; JOINED.
DR      EMBL; Y19131; CAB65555.1; JOINED.
DR      EMBL; Y19132; CAB65555.1; JOINED.
DR      EMBL; Y19133; CAB65555.1; JOINED.
DR      EMBL; Y19134; CAB65555.1; JOINED.
DR      EMBL; Y19135; CAB65555.1; JOINED.
DR      EMBL; Y19136; CAB65555.1; JOINED.
DR      EMBL; Y19137; CAB65555.1; JOINED.
DR      EMBL; Y19138; CAB65555.1; JOINED.
DR      EMBL; Y19139; CAB65555.1; JOINED.
DR      EMBL; Y19140; CAB65555.1; JOINED.
DR      EMBL; Y19141; CAB65555.1; JOINED.
DR      EMBL; Y19142; CAB65555.1; JOINED.
DR      EMBL; Y19143; CAB65555.1; JOINED.
DR      EMBL; AL163218; CAB90392.1; -
DR      EMBL; AL163217; CAB90389.1; -
DR      PIR; A56318; A56318.
DR      HSP; P00763; LDPO.
DR      MEROPS; S01.156; -.
DR      Genew; HGNC:9490; PRSS7.
DR      MIM; 606635; -.
DR      MIM; 226200; -.
DR      GO; GO:0005903; C:brush border; TAS.
DR      InterPro; IPR000859; CUB.
DR      InterPro; IPR009003; Cys Ser trypsin.
DR      InterPro; IPR002172; LDL_receptor_A.
DR      InterPro; IPR000998; MAM_domain.
DR      InterPro; IPR001254; Peptidase S1.
DR      InterPro; IPR001314; Peptidase S1A.
DR      InterPro; IPR000082; SEA_domain.
DR      InterPro; IPR001190; Srcr_receptor.
DR      Pfam; PF00431; CUB; 2.
DR      Pfam; PF00057; ldl_recept_a; 2.
DR      Pfam; PF00629; MAM; 1.
DR      Pfam; PF01390; SEA; 1.
DR      Pfam; PF05330; SRCR; 1.
DR      Pfam; PF00089; trypsin; 1.
DR      PRINTS; PR00722; CHYMOTRYPSIN.
DR      PRINTS; PR00261; LDLRECEPTOR.
DR      PRINTS; PR00020; MAMDOMAIN.
DR      SMART; SM00042; CUB; 2.
DR      SMART; SM00192; LDLa; 2.
DR      SMART; SM00137; MAM; 1.
DR      SMART; SM00200; SEA; 1.
DR      SMART; SM00202; SR; 1.
DR      SMART; SM00020; TYP_SPC; 1.
DR      PROSITE; PS01180; CUB; 2.
DR      PROSITE; PS01209; LDLRA_1; 2.
DR      PROSITE; PS00068; LDLRA_2; 2.
DR      PROSITE; PS00740; MAM_1; 1.
DR      PROSITE; PS00060; MAM_2; 1.
DR      PROSITE; PS00024; SEA; 1.
DR      PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR      PROSITE; PS02287; SRCR_2; 1.
DR      PROSITE; PS02040; TRYPSIN_DOM; 1.
DR      PROSITE; PS00134; TRYPSIN_HIS; 1.
DR      PROSITE; PS00135; TRYPSIN_SER; 1.
KW      Signal-anchor; Glycoprotein; Myristate; Serine protease;
KW      Zymogen; Transmembrane; Repeat; Lipoprotein.
FT      CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT      CHAIN 785 1019 CATALYTIC CHAIN (LIGHT CHAIN).
FT      DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT      (POTENTIAL).
FT      DOMAIN 48 1019 EXTRACELLULAR (POTENTIAL).

```

```

FT      DOMAIN 52 169 SEA.
FT      DOMAIN 182 223 LDL-RECEPTOR CLASS A 1.
FT      DOMAIN 225 334 CUB 1.
FT      DOMAIN 342 504 MAM.
FT      DOMAIN 524 634 CUB 2.
FT      DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.
FT      DOMAIN 678 771 SRCR.
FT      DOMAIN 785 1019 SERINE PROTEASE.
FT      ACT SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT SITE 971 971 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT SITE 971 971 N-myristoyl glycine (Potential).
FT      LIPID 2 2 BY SIMILARITY.
FT      DISULFID 184 197 BY SIMILARITY.
FT      DISULFID 191 210 BY SIMILARITY.
FT      DISULFID 204 221 BY SIMILARITY.
FT      DISULFID 643 655 BY SIMILARITY.
FT      DISULFID 650 668 BY SIMILARITY.
FT      DISULFID 662 677 BY SIMILARITY.
FT      DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
FT      DISULFID 810 826 BY SIMILARITY.
FT      DISULFID 910 977 BY SIMILARITY.
FT      DISULFID 941 956 BY SIMILARITY.
FT      DISULFID 967 995 BY SIMILARITY.
FT      CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 179 179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 335 335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 388 388 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 440 440 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 503 503 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 630 630 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 682 682 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 706 706 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 848 848 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 887 887 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 909 909 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CARBOHYD 949 949 N-LINKED (GLCNAC. .) (POTENTIAL).
FT      CONFLICT 134 134 Q -> E (IN REF. 3).
FT      CONFLICT 732 732 S -> P (IN REF. 3).
FT      CONFLICT 754 771 SQQLQSLRLQCNHKS -> RRNAKNEIDALSPILIA
      (IN REF. 3).

```

```

Query Match 24.9%; Score 582.5; DB 1; Length 1019;
Best Local Similarity 33.8%; Pred. No. 3.9e-42;
Matches 135; Conservative 62; Mismatches 132; Indels 71; Gaps 15;

Qy 67 HF-----IPRKQLCDGELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATG 119
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 647 HFQCKNGECVPLVNLCDGHLHCEGDSDEADCVRF-----NGTTN 686

Qy 120 N-----WFSACFDNFTALAEATACRQMGY-----SSKPTFRAVIGPPQDLDDVE 154
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 687 NNGLVRFRIQSIWHTACAEWITQISNDYCOLLGLSGNSKPIF-STDGGPFVKLTAP 745

Qy 165 ----ITENSQELMRNNSGPGSLVSLHC--LACGSKL-----TPRVVGGSEASVDSW 214
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 746 DGHLLITPSQO-----CLQDSILRLQCNHKSCKKLAQDITPKLVGGSNAKEGAW 796

Qy 215 PMQVSIQYDKQHVCGSILDPHWLVTAHC-FRKHTDVFNWKVRAGSKLGSPPSLAVAK 273
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 797 PMVVGLYGGRLLCGASLVSSDWLVSAAHCYVGRNLEPSKWTAILGLHMKSNLTSPQTVP 856

Qy 274 IIT--TEFPMY---PKNDIALMKLOPLTFTSGTVRPICLPFFDELTLPATLWIGWG 328
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 857 RLIDEIVPHNRRKNDIAVMHLEFKVNYTDYIQLPCLPEENVQVFPGRNCSAGWG 916

Qy 329 FTKQNGKMSDILLQASVQVIDSTRCNADDAVOGEVTEKMCAGIEGGVDTCCQSGSGP 388
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 917 -TVVYQTTANILQEAADVLLSNRCQ-QQMPEYNITENMICAGYEKGIDSCQSGSGP 974

```


QY 249 TDVFN-WKVRAGSKLGSFSLAVAKIIIIIBENMYP-----KONDIALMKLQPLPT 299
 Db 228 NRVLNRWRFAGAVARTSPHAFVQLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSSLP 287
 QY 300 PSGTVRPTCLPFFDELPATPLMIIGWFTKONGKMSDILLOASVOVIDSTRCNADDA 359
 Db 288 LTEYIQVCLPAAGALVDGKVCVTGWENT-QYGGQAMVLOARPIISNEVCNPSDF 346
 QY 360 YQGEVTERKMCAGIPEGGVDFRCQSGSGPLMYQ-----SDQWVVGVISWYGGCGSPST 414
 Db 347 YGNQIKPMFCAGYPEGGIDACQSGSGPFVCEDSISGTSWRLCIGVSWGTGCGALARKP 406
 QY 415 GYTKVSAVLANWYNVK 432
 Db 407 GYTKVDFREWFKAIK 424

RESULT 11

TMSS5_HUMAN
 ID TMSS5_HUMAN STANDARD; PRT; 457 AA.
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 40, Last sequence update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMPS55.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741986;
 RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 RT "Spinesin"/TMPS5, a novel transmembrane serine protease, cloned from
 RL human spinal cord.";
 RJ J. Biol. Chem. 277:6806-6812(2002).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 CC neurons, in their axons, and at the synapses of motoneurons in the
 CC spinal cord.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB028140; BAB20375.1; --
 DR HSPF; P00763; IDPO.
 DR Genew; HGNC:14908; TMPS55.
 DR MIM; 606751; --
 DR MEROPS; S01.313; --
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp SPc; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; FALSE_NEG.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 71 457 (POTENTIAL).
 FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 218 207 SRCR.
 FT ACT_SITE 258 258 SERINE PROTEASE.
 FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 217 218 CLEAVAGE (POTENTIAL).
 FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC... (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC... (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC... (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC... (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC... (POTENTIAL).
 SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2851 CRC64;
 Query Match 24.5%; Score 574; DB 1; Length 457;
 Best Local Similarity 32.2%; Pred. No. 7.2e-42;
 Matches 129; Conservative 70; Mismatches 151; Indels 50; Gaps 12;
 QY 60 FLC---GQPLHFTPRKQLCDGE--LDCPLGEDDEHCVKSPFPEGPVAVRLSKDRSTLQVL 114
 Db 71 YLCPAASQPI-----SGTLQDEEITLSCSASEEALLPALPK--TVSFRLNSDFLLEAQ 124
 QY 115 DSATGNWFSACFDNFTAEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQE--- 171
 Db 125 VRQPRWLLVCHGEGWSPALGLQICWSLGH-----LRLTHKGVNLTDLKLNSSQEFA 176
 QY 172 -----LRMNSSGPCLSGLSVLSLHCLACKSLKTPRVVGGSEASVDSPWQV 218
 Db 177 QLSFRLGGLFEAEAWQPRNN--CTSGQVWSLRCSECGARPLASRIVGGOSVAFGRWPQA 233
 QY 219 SIQYDKQHVCGGSIIDPHWVLTAAHCFR--KH'DVFNWVKRAGSDKLGSPSPSLAVAKIII 276
 Db 234 SVALGRFHTCGGSVLAPRWVVTAAHCHWSFRLARLSSWRVHAG--LVSHSAVRPQGAL 290
 QY 277 IEF---NPMYPKDN---DIALMKLQPLFTFSQVTRPICLPFFDELTPTATPLMIWGFT 330
 Db 291 VEIIIPLYSAQNDYDVALRLQALNFSDTVGAVCLPAKEHQHPKSGRCWVSWGHT 350
 QY 331 KQNGKMSDILLQASVOVIDSTECNADDAVQGEVTERKMCAGIPEGGVDFRCQSGSGPLM 390
 Db 351 HPSTYSSDMQLQDTVVPLSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQSGSGPLV 410
 QY 391 Y-QSDQWVVGVISWYGGCGSPSTPGVYTKVSAVLANWYN 429
 Db 411 CPDGDTRWLVGVVSWGRACAEPNHGPYAKVAEFLDWIHD 450

RESULT 12

ENTK_BOVIN
 ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
 AC P98072;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTK.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Ductenium;
 RX MEDLINE=94329561; PubMed=8052624;


```
FT CONFLICT 808 808 R -> Y (IN REF. 3)
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

Query Match 24.4%; Score 571; DB 1; Length 1035;
Best Local Similarity 35.0%; Pred. No. 3.9e-41;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

QY 59 IPRKCLDGLDCLPGEDEEHCXKSFPEGPVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
DB 672 IPLVNLCDGPHCKDGSDEAHCVRLLF-NGTTDSSGLVQFR--IQSI-----WHVACAEN 722
QY 129 FTEALAEACRMGY-----SSKPTFRFAVEIGPDODLDVVE-----ITENSQBLMRNNSGP 180
DB 723 WTTQISDDVCQLGLGTCNNSVPITF-STGGGPPYVNLNTPNGSLILPSSQ-----772
QY 131 CLSGSLVSLHC--LACGSLKT-----PRVVGEEASVDSWPQVSIQYDKHVCVGGSLD 234
DB 773 CLEDSLILQLQNYKSCGKKLVQEVSPKIVGSDSREGAWPWVVALYFDQDQVCGASLVS 832
QY 235 PFWVLTAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND 288
DB 833 RDLVLSAAHCYGRNWEPSKWKAVLGLHMASNLTSPOIETRLDIQIVINPHYKRNND 892
QY 289 IALMKLQPLTFSGTVRPICLPFPDEELTPATPLWIGWGTGKONGKMSDILLQASVQV 348
DB 893 IAMHLEMKVNYDTYIQIPICLPEENQVPPGRICSIAGGALIIYQ-GSTADVLQEAADVPL 951
QY 349 IDSFTCNADDAVGGVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ--SDQVHVGVISVGYG 407
DB 952 LSNEKQ--QMPYENITENWVCAGYEAGVDSGCCDSGGPLMCCENRKLWLLAGVTSFYQ 1010
QY 408 CGGPSTPGVYTKVSAYLWNI 427
DB 1011 CALPNRPGVYARVPRTEWI 1030

RESULT 13
HEPS_RAT STANDARD; PRT; 416 AA.
AC Q05511;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-).
GN HPN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93305733; PubMed=8318546;
RA Farley D., Raymond F., Nick H.;
RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
ET proteinase.";
RL Biochim. Biophys. Acta 1173:350-352(1993).
CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
CC of cell morphology.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X70900; CAA50256.1; -
DR PIR; S33777; S33777.
DR HSSP; P00763; 1DPO.
```

```
DR MEROPS; S01.224; -
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Src receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; trypt; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161 SERINE PROTEASE HEPsin, NON-CATALYTIC
FT CHAIN (POTENTIAL) CHAIN (POTENTIAL)
FT CHAIN 162 416 SERINE PROTEASE HEPsin, CATALYTIC CHAIN
FT CHAIN (POTENTIAL) CHAIN (POTENTIAL)
FT CHAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT CHAIN 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CHAIN (POTENTIAL) (POTENTIAL).
FT CHAIN 44 416 EXTRACELLULAR (POTENTIAL).
FT CHAIN 162 416 SERINE PROTEASE
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
FT DISULFID 187 203 BY SIMILARITY.
FT DISULFID 321 337 BY SIMILARITY.
FT DISULFID 348 380 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 416 AA; 44926 MW; E5A9F8FA950E180 CRC64;

Query Match 24.3%; Score 568.5; DB 1; Length 416;
Best Local Similarity 30.5%; Pred. No. 1.9e-41;
Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;

QY 19 KPRIPMETRKVGIPIIIIALLSLASIIIVVLIKVILDKYFLCGPLHPIPRKQLDGE 78
DB 14 RPKVAALT--VGLTLFTIGAGASWAIIVTLR-----44
QY 79 LDCPLGEDEEHCXKSFPEGPVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALAEATC 138
DB 45 -----SDQE-----PLYQVQLSPGDSRLVLVDKTEGTWLLCSSRSNARVAGLGC 89
QY 139 ROMGY-----SSKPTFRFAVEIG-----PDQLDVVEITENSQBLMRNNS 178
DB 90 EEMGFLRALAHSELVDVRTAGANGTSGFCVDEGGLPLAQRLLDVIGVCD-----138
QY 179 GPLCSGLSVSLHCLAGC-KSLKTPRVVVGEEASVDSWPQVSIQYDKHVCVGGSLDPHW 237
DB 139 --CFRGRFLTATCCDCGRRLKLPVDRIVGGDSGLGWPQVSLRYDGLTLCGSGSLSDGW 196
QY 238 VLTAACHCFRRKHTDVFN-WKVRAGSDKLGSPPSLAVAKIIIEFNPMYP-----KDND 288
DB 197 VLTAACHCFERNVLSRWRVFAVARTSPHVAQLGVQAVIYHGGYLPPRDPIDENSND 256
QY 289 IALMKLQPLTFSGTVRPICLPFPDEELTPATPLWIGWGTGKONGKMSDILLQASVQV 348
DB 257 IALVHLSSSLPTEYIQPVCLPAAGQALVDGKVTVTGNGT-QFYQQAVVLQEARVPI 315
QY 349 IDSTRCNADDAVQCEVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ-----SDQVHVGVIS 403
DB 316 ISNEVCNSPDFYGNQIKPKMFCAGYPEGGLDAGCGSGGHFVCEDRISGTSRWLCGIVS 375
QY 404 WGVCGGPGSTPGVYTKVSAYLWNIYVNWK 432
DB 376 WGTGTCALARKPGVYTKVIDPFWIFQAIA 404

RESULT 14
HATT_HUMAN STANDARD; PRT; 418 AA.
ID HATT_HUMAN
AC O60235;
```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Airway trypsin-like protease precursor (EC 3.4.21.-).
 GN HAT.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=98234382; PubMed=9565616;
 RA Yamaoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yamaoka S.,
 RA "Cloning and characterization of the cDNA for human airway trypsin-
 RT like protease.";
 RL J. Biol. Chem. 273:11895-11901(1998).
 RN [2]
 RN SEQUENCE OF 187-206, AND CHARACTERIZATION.
 RX MEDLINE=97224034; PubMed=9070615;
 RA Yamaoka S., Onishi T., Kawano S., Teuchihashi S., Ogawara M.,
 RA Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;
 RT "Purification, characterization, and localization of a novel
 RT trypsin-like protease found in the human airway.";
 RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).
 CC -!- FUNCTION: May play some biological role in the host defense system
 CC on the mucous membrane independently of or in cooperation with
 CC other substances in airway mucous or bronchial secretions.
 CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of
 CC arginine residues at the pi position of certain peptides, cleaving
 CC Boc-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and
 CC having an optimum pH of 8.5 with this substrate.
 CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl
 CC fluorophosphate, leupeptin, antipain, aprotinin, and soybean
 CC trypsin inhibitor, but hardly inhibited by secretory leukocyte
 CC protease inhibitor at 10 microm.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
 CC cleavage and secreted.
 CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous
 CC glands of the bronchi and trachea.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 SEA domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; AB002134; BAA28691.1; --
 DR HSSP; P00750; IRTF.
 DR MEROPS; S01.301; --
 DR MIM; 605369; --
 DR GO; GO:0005576; C:extracellular; TAS.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008233; F:peptidase activity; TAS.
 DR GO; GO:0007585; P:respiratory gaseous exchange; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS00024; SEA; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW

Glycoprotein. 1 185 AIRWAY TRYPSIN-LIKE PROTEASE, NON-
 CHAIN CATALYTIC CHAIN.
 FT CHAIN 187 418 AIRWAY TRYPSIN-LIKE PROTEASE, CATALYTIC
 FT CHAIN CHAIN.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 41 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 42 418 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 44 164 SEA.
 FT DOMAIN 187 417 SERINE PROTEASE.
 FT ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 173 292 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 212 228 BY SIMILARITY.
 FT DISULFID 337 353 BY SIMILARITY.
 FT DISULFID 364 393 BY SIMILARITY.
 FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 418 AA; 46263 MW; F4BC1DB020CFBBD0 CRC64;

Query Match 24.3%; Score 568; DB 1; Length 418;
 Best Local Similarity 32.4%; Pred. No. 2.1e-41;
 Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIALLSLASIIIVVLKVIDKYFLCGQPLHPIRK--OLCDGELDCPL----- 83
 DB 16 FYVVCIVVAGVILAVTIALIV---YFLAFDQKSYFYRSSFOLLNVEYNSQLNSPATQE 72
 QY 84 -----GEDEEHCVKSPPEG-----PAVAVRLSKD---RSTLQVLDSATGNWFSACFD 127
 DB 73 YRTLSGRISLTKTKESNLRNQFIKRAHVAKLRQDGSVGRADVVMKQFTENNGASM- 131
 QY 128 NFEALAEATACROMGYSSRPTPRAVEIGDQDLVDVVEITENSQELMRN--SSGCLSGS 185
 DB 132 ---KSRIEVLRLQMLNNS---GNLEINP---STEITSLTDQAAANWLINCEGAGPDLI-- 180
 QY 186 LVSLHCLACGSKLTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSILDPHWLTAACHF 245
 DB 181 -----TLSEQRILGTEAEBSWPMQVSLRLNNAHCGGSLINNMILTAACHF 229
 QY 246 RKHTDVENKVRAGSKLGSFPSSL--AVAKIIIEFNPMYKNDIALMKLOFPFTFSCT 303
 DB 230 RSNNSPRDMATSGIST--TFPKLRMRVRNLIHNHYKSATHENDIALVRLNSVTFTKD 287
 QY 304 VRPICLPFDFDEELTPATPLWILGWFTKQNGGMSDILLQASVQVIDSTRCNADDAYQGE 363
 DB 288 IHSVCLPAATQNIIPPGSTAYVTGWAQAEYAGHTVPE-LRQGGVRIISNDVCMAPHSYNGA 346
 QY 364 VTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQ--WHVVGIVSWGYGCGGFPSTPGVYTKVS 421
 DB 347 ILSGMLCAGVPGQGVDDACCGDSGGPLVQEDSRRLWFIWGVISWGDQCGLPDKPGVYTRVT 406
 QY 422 AYLNWI 427
 DB 407 AYLDWI 412

RESULT 15
 ENTX_MOUSE
 ID ENTX_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enteropeptidase (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Duoenum;
 RX MEDLINE=98147142; PubMed=9486188;
 RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.:
 RT "Structure of murine enterokinase (enteropeptidase) and expression in
 RL small intestine during development.";
 RL Am. J. Physiol. 274:G342-G349(1998).
 CC -!- FUNCTION: Responsible for initiating activation of pancreatic
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in
 CC turn activates other proenzymes including chymotrypsinogen,
 CC procarboxypeptidases, and proelastases (By similarity).
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
 CC trypsinogen.
 CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
 CC multidomain (heavy) chain linked by a disulfide bond (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 2 CUB domains.
 CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 1 MAM domain.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U73378; AAB37317.1; -.
 CC HSP: Q07954; 1CR8.
 CC MEROPS: S01.156; -.
 CC MGD: MG1.1197523; Prss7.
 CC InterPro: IPR000859; CUB.
 CC InterPro: IPR009003; Cys_Ser_trypsin.
 CC InterPro: IPR002172; LDL_receptor_A.
 CC InterPro: IPR000998; MAM_domain.
 CC InterPro: IPR001254; Peptidase_S1.
 CC InterPro: IPR001314; Peptidase_S1A.
 CC InterPro: IPR009020; Protease_inhib.
 CC InterPro: IPR000082; SEA_domain.
 CC InterPro: IPR001190; Srcr_receptor.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF00057; ldl_recept_a; 2.
 CC Pfam: PF00629; MAM; 1.
 CC Pfam: PF01390; SEA; 1.
 CC Pfam: PF00530; SRCR; 1.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC PRINTS: PR00261; LDLRECEPTOR.
 CC PRINTS: PR00020; MAMDOMAIN.
 CC SMART: SM00042; CUB; 2.
 CC SMART: SM00192; LDLA; 2.
 CC SMART: SM00137; MAM; 1.
 CC SMART: SM00200; SEA; 1.
 CC SMART: SM00202; SR; 1.
 CC SMART: SM00020; Tryp_SPC; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS01209; LDLA; 1; 2.
 CC PROSITE: PS00687; LDLA; 2; 2.
 CC PROSITE: PS00740; MAM; 1; 1.
 CC PROSITE: PS00650; MAM; 2; 1.
 CC PROSITE: PS00024; SEA; 1.
 CC PROSITE: PS00420; SRCR; 1; FALSE_NEG.
 CC PROSITE: PS0287; SRCR; 2; 1.
 CC PROSITE: PS0240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.

KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
 KW Transmembrane; Repeat.
 FT CHAIN 1 839 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 48 1069 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 52 169 SEA.
 FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 270 379 CUB 1.
 FT DOMAIN 387 549 MAM.
 FT DOMAIN 569 679 CUB 2.
 FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 723 816 SRCR.
 FT DOMAIN 830 1069 SERINE PROTEASE.
 FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 229 242 BY SIMILARITY.
 FT DISULFID 236 255 BY SIMILARITY.
 FT DISULFID 249 266 BY SIMILARITY.
 FT DISULFID 688 700 BY SIMILARITY.
 FT DISULFID 695 713 BY SIMILARITY.
 FT DISULFID 707 722 BY SIMILARITY.
 FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 859 875 BY SIMILARITY.
 FT DISULFID 959 1027 BY SIMILARITY.
 FT DISULFID 991 1006 BY SIMILARITY.
 FT DISULFID 1017 1045 BY SIMILARITY.
 FT CARBOHYD 147 147 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 197 197 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 380 380 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 579 579 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 727 727 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 770 770 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 791 791 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 897 897 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 936 936 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 999 999 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1069 AA; 118735 MW; E62549B463743CD CRC64;
 Query Match 23.6%; Score 552.5; DB 1; Length 1069;
 Best Local Similarity 32.5%; Pred. No. 1.6e-39;
 Matches 124; Conservative 71; Mismatches 150; Indels 37; Gaps 12;
 QY 69 IPRKQICDGLDCLGEDEHCVKSPFEPFAVRLSKDRSTQVLVD-SATGNWFACFD 127
 Db 701 IPLGNLCSYPHCRDGSDEASCVRF-----LNGTRSNNGVLQFNIHSHWACAE 750
 QY 128 NFEALAEATACROMGYSSKPTFRAVBI---GPDQDLVWEITENSQELMRNRSFGCLSG 184
 Db 751 NWTQISNEVCHLLGLGSANSPISSTGGP-----FVRVQAQNGSLILTSLQCSQD 805
 QY 185 SLVSLHC--LACGKSLKT----PRVVGGEASVDSWPMQVSIQYDKQH-----VCGSGILD 234
 Db 806 SLILLQCNHKSCEKKVTQKVSFKVGGSDAQAGAPWVVALYHRDRSTDRLLCGASLVS 865
 QY 235 PHVVLTAHC-FRKHTDVFNNKVRAGSKLGSPPSLAVAKILII--TEFNPMYP---KDND 288
 Db 866 SDMLVSAACHVCYARNLDPTRWTAVLGHMQSNLTSFQVVRVVDQIVINPHYDRRKVND 925
 QY 289 IALMKIQPLTFSGTVRPICLPFFDEBELTPATPLWIGMGFTKQNGGKMSDILLQASVQV 348
 Db 926 IAMHLEFKVNYTDYIQICLPEENQIFIGRTCSIAGWGYDKINAGSTVDVLKEADVPL 985
 QY 349 IDSTRCNADDAVQGEVTERKMKAGIEGGVDVTCGSGSGPLMYQ-SDQWHVGVISWGYG 407

Db 986 ISNEXCQ-QQLPEYNITESMICAGYEEGGIDSCQGDSCGGLMCOENNEWFLVGTSGVQ 1044
QY 408 CCGPSTPGVYTKVSAYLWYN 429
Db 1045 CALPNHPGVYVRSQFIEWHS 1066

Search completed: June 7, 2004, 08:39:20
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 45 Seconds
(without alignments)
3050.009 Million cell updates/sec

Title: US-10-030-688-2

Perfect score: 2342
Sequence: 1 MDPDSDFPLNSLDVKPLRKP.....VYTKVSAYLWNVNWKAEL 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:**

- 1: sp:archaea:**
- 2: sp:bacteria:**
- 3: sp:fungi:**
- 4: sp:human:**
- 5: sp:invertebrate:**
- 6: sp:mammal:**
- 7: sp:mhc:**
- 8: sp:organelle:**
- 9: sp:phase:**
- 10: sp:plant:**
- 11: sp:rodent:**
- 12: sp:virus:**
- 13: sp:vertebrate:**
- 14: sp:unclassified:**
- 15: sp:virus:**
- 16: sp:bacteriap:**
- 17: sp:archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2167	92.5	405	Q96E86	Q96E86 homo sapien
2	704.5	30.1	471	Q8CFE0	Q8CFE0 mus musculus
3	695.5	29.7	581	Q8BYE2	Q8BYE2 homo sapien
4	688	29.4	453	Q8LZA6	Q8LZA6 mus musculus
5	686.5	29.3	537	Q8BYE1	Q8BYE1 homo sapien
6	686.5	29.3	558	Q8CYM4	Q8CYM4 homo sapien
7	686.5	29.3	767	Q9DGR2	Q9DGR2 xenopus lae
8	670.5	28.6	492	Q86TV3	Q86TV3 homo sapien
9	648	27.7	490	Q7TND4	Q7TND4 mus musculus
10	646	27.6	490	Q920K3	Q920K3 rattus norv
11	636.5	27.2	445	Q8CJL7	Q8CJL7 rattus norv
12	631.5	27.0	455	Q8CDP0	Q8CDP0 mus musculus
13	612	26.1	371	Q8CJL6	Q8CJL6 rattus norv
14	545	23.3	777	Q8CJL9	Q8CJL9 mus musculus
15	536	22.9	417	Q8VHJ4	Q8VHJ4 rattus norv
16	527	22.5	279	Q9QZ74	Q9QZ74 rattus norv

17	523	22.3	417	11	Q8VDV1	Q8VDV1 mus musculus
18	523	22.3	417	11	Q8VHK8	Q8VHK8 mus musculus
19	521.5	22.3	279	11	Q7TNX3	Q7TNX3 mus musculus
20	510	21.8	328	11	Q8OZ40	Q8OZ40 rattus norv
21	504.5	21.5	329	6	Q9GL10	Q9GL10 ovis aries
22	504.5	21.5	1524	13	Q91674	Q91674 xenopus lae
23	503.5	21.5	336	11	Q8OYD8	Q8OYD8 mus musculus
24	502	21.4	638	11	Q8ROP5	Q8ROP5 mus musculus
25	501.5	21.4	439	11	Q8BHM9	Q8BHM9 mus musculus
26	501.5	21.4	556	13	Q8O3D5	Q8O3D5 brachydanio
27	500.5	21.4	1111	11	Q8OYN4	Q8OYN4 rattus norv
28	499.5	21.3	327	4	Q8N171	Q8N171 homo sapien
29	499.5	21.3	572	11	Q8BIK6	Q8BIK6 mus musculus
30	498.5	21.3	377	6	P79343	P79343 bos taurus
31	498.5	21.3	417	11	Q8BZ10	Q8BZ10 mus musculus
32	496	21.2	643	6	Q97506	Q97506 sus scrofa
33	493.5	21.1	421	11	Q6O491	Q6O491 cavia porce
34	493	21.1	416	4	Q8GT26	Q8GT26 homo sapien
35	488	20.8	423	11	Q8BM10	Q8BM10 mus musculus
36	488	20.8	812	11	Q9ROW3	Q9ROW3 rattus norv
37	486.5	20.8	416	11	Q8BZ30	Q8BZ30 mus musculus
38	486.5	20.8	416	11	Q8BZ13	Q8BZ13 mus musculus
39	486	20.8	389	13	Q9PVX7	Q9PVX7 xenopus lae
40	485.5	20.7	326	13	Q7ZZ80	Q7ZZ80 brachydanio
41	482.5	20.6	284	4	Q8NFS6	Q8NFS6 homo sapien
42	481	20.5	321	4	Q96RZ8	Q96RZ8 homo sapien
43	479.5	20.5	845	13	Q9DGR1	Q9DGR1 xenopus lae
44	478.5	20.4	415	6	Q29015	Q29015 sus sp. pre
45	477	20.4	1379	5	Q9V4N6	Q9V4N6 drosophila

ALIGNMENTS

RESULT 1
Q96E86
ID Q96E86 PRELIMINARY; PRT; 405 AA.
AC Q96E86
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -
DR HSSP; P00761; IAN1.
DR MEROPS; S01.034; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR090003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMC0192; LDLa; 1.
DR SMART; SMC0202; SR; 1.
DR SMART; SMC0020; Tryp_Spc; 1.
DR PROSITE; PS50287; SRCF_2; 1.

DR PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

FW Hydrolase; Protease; Serine protease; Transmembrane.

FT NON_TER 1

SQ SEQUENCE 405 AA; 44474 MW; 951ACD529D48E04 CRC64;

Query Match 92.5%; Score 2167; DB 4; Length 405;

Best Local Similarity 100.0%; Pred. No. 8.3e-202;

Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 IIALSLASIIIVVLLKVLIDKYFLCGPLHPIPRKQCDGELDCPLGEDSEHCVK 93

DB 4 IIALSLASIIIVVLLKVLIDKYFLCGPLHPIPRKQCDGELDCPLGEDSEHCVK 63

QY 94 PEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAEACRQGYSSKPTFRAVE 153

DB 64 PEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAEACRQGYSSKPTFRAVE 123

QY 154 IGPDDLDVWEITENSQELRMNSGPGCLSGSLVSLHCLACGKSLKTPRVVGGSEASVDS 213

DB 124 IGPDDLDVWEITENSQELRMNSGPGCLSGSLVSLHCLACGKSLKTPRVVGGSEASVDS 183

QY 214 WFWQVSIQYDKHGVCGGSLDPHWLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 273

DB 184 WFWQVSIQYDKHGVCGGSLDPHWLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 243

QY 274 IIIIEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDEBELPATPLWIIGWFTKQN 333

DB 244 IIIIEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDEBELPATPLWIIGWFTKQN 303

QY 334 GSKMSDILLOASVQVIDTRCNADDAVAGEVTEKMWKAGIPEGVDTCQDGGGGLMYQS 393

DB 304 GSKMSDILLOASVQVIDTRCNADDAVAGEVTEKMWKAGIPEGVDTCQDGGGGLMYQS 363

QY 394 DQWVVGVISVWGCGGPGSTPGVYTKVSAYLNWYVWKAEL 435

DB 364 DQWVVGVISVWGCGGPGSTPGVYTKVSAYLNWYVWKAEL 405

RESULT 2

Q8CFEO

ID Q8CFEO PRELIMINARY; PRT; 471 AA.

AC Q8CFEO;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Similar to mosaic serine protease (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]_

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Breast tumor;

RA Strausberg R.;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC042878; AAH42878.1; -

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0004263; F:chymotrypsin activity; IEA.

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0005044; F:scavenger receptor activity; IEA.

DR GO; GO:0004295; F:trypsin activity; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR002172; Ldl receptor A.

DR InterPro; IPR001254; Peptidase S1_

DR InterPro; IPR001314; Peptidase_S1A.

DR InterPro; IPR001190; Srcr_receptor.

DR Pfam; PF00057; ldl_recept_a; 1.

DR Pfam; PF00530; SRCR; 1.

DR Pfam; PF00089; trypsin; 1.

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR SMART; SM00202; SR; 1.

DR SMART; SM00020; Tryp_SPC; 1.

DR PROSITE; PS00287; SRCR_2; 1.

DR PROSITE; PS00240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

FW Protease.

FT NON_TER 1

SQ SEQUENCE 471 AA; 52535 MW; ED56CFB6B7C3BCC4 CRC64;

Query Match 30.1%; Score 704.5; DB 11; Length 471;

Best Local Similarity 36.2%; Pred. No. 1.5e-59;

Matches 153; Conservative 70; Mismatches 167; Indels 33; Gaps 9;

QY 25 ETPRKVGIPIIIALSLASIIIVVLLKVLIDKYFLCG-----QPLHFIPTKQL-CD 76

DB 65 ETPRKVGIPIIIALSLASIIIVVLLKVLIDKYFLCG-----QPLHFIPTKQL-CD 76

QY 77 GELDCPLGEDSEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAE 136

DB 117 GELDCPLGEDSEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAE 136

QY 137 ACRQGYSSKPTFRAVEIGPDODLDVWEITENSQELRMNSGPGCLSGSLVSLHCLACGK 196

DB 166 ACRQGYSSKPTFRAVEIGPDODLDVWEITENSQELRMNSGPGCLSGSLVSLHCLACGK 196

QY 197 SLKTPRVVGGSEASVDSWFWQVSIQYDKHGVCGGSLDPHWLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 254

DB 224 SLKTPRVVGGSEASVDSWFWQVSIQYDKHGVCGGSLDPHWLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 254

QY 255 KVRAGSDKLGSFSLAVAKIIIEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDEBELPATPLWIIGWFTKQN 313

DB 284 KVRAGSDKLGSFSLAVAKIIIEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDEBELPATPLWIIGWFTKQN 313

QY 314 BELTPATPLWIIGWFTKQNKGKMSDILLOASVQVIDTRCNADDAVAGEVTEKMWKAGIPEGVDTCQDGGGGLMYQS 373

DB 344 BELTPATPLWIIGWFTKQNKGKMSDILLOASVQVIDTRCNADDAVAGEVTEKMWKAGIPEGVDTCQDGGGGLMYQS 373

QY 374 PEGGVDTCQDGGGGLMYQSDQWVVGVISVWGCGGPGSTPGVYTKVSAYLNWYVWKAEL 432

DB 404 PEGGVDTCQDGGGGLMYQSDQWVVGVISVWGCGGPGSTPGVYTKVSAYLNWYVWKAEL 432

QY 433 AEL 435

DB 464 SEV 466

RESULT 3

Q8BYE2

ID Q8BYE2 PRELIMINARY; PRT; 581 AA.

AC Q8BYE2;

DT 01-JUN-2001 (TREMELrel. 17, Created)

DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Membrane-type mosaic serine protease.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]_

RP SEQUENCE FROM N.A.

RC TISSUE=lung;

RX MEDLINE=21167333; PubMed=11267681;

RA Kim D.R., Sharmir S., Inoue M., Kido H.;

RT "Cloning and expression of novel mosaic serine proteases with and

RL without a transmembrane domain from human lung."

RL Biochim. Biophys. Acta 1518:204-209 (2001).

DR EMBL; AB048796; BAB39741.1; -

DR HSSP; P00763; IDPO.

DR MEROPS; S01.087; -

DR GO; GO:0016021; C:integral to membrane; NAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR	InterPro; IPR0093003; Cys_Ser_trypsin.
DR	InterPro; IPR002172; LDL_receptor A.
DR	InterPro; IPR001254; Peptidase S1.
DR	InterPro; IPR001314; Peptidase S1A.
DR	InterPro; IPR001190; Srcr receptor.
DR	Pfam; PF00057; ldl_recept_a; 1.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00192; LDla; 1.
DR	SMART; SM00202; SR; 1.
DR	SMART; SM00020; Tryp_Spc; 1.
DR	PROSITE; PS02087; SERC_2; 1.
DR	PROSITE; PS02040; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Hydrolase; Protease; Serine protease.
SQ	SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA4A4 CRC64;

Query Match	29.7%; Score 595.5; DB 4; Length 581;
Best Local Similarity	35.4%; Pred. No. 1.5e-58;
Matches 152; Conservative 69; Mismatches 161; Indels 47; Gaps 11;	

QY	22	IPMETFR--KVGIPFIITALLSLAGIIIVWLKVILDKYVFLCGQ-----PLH 67
DB	149	LPKFTWREGQQLPGLICGVLLIILVLSLIIL-----FQWQHTGIRHKQEESCPEKH 202
QY	68	FIPRKQLCDGELDCPLGEDEHECHVKSPPEGPAVARLRSKORSTIQVLDSATGNWFSA CFD 127
DB	203	AV-----RCDGWVDCKLKSDELGC-----VRFDWDKSLKIYSGSSHQWLPICSS 247
QY	128	NFTALAEATACROMGYSKPTFRAVEIGPDQDLDVVEITENSQELRWNSGGCLSGSLV 187
DB	248	NWNDSYSSEKTCRQGTFES--AHRTEVAHRDFANSPFILRNSTIQBSLHRSHCFPSORI 305
QY	188	SLHCLACKSLKTPRVVGGEEASVDSWPQVSIQYDKHQVCGGSILDPHWLTAAHCF-- 245
DB	306	SLQCSHGCLRAMTGRIVGGALASDKPWQVSLHFGTTHICGTLIIAQWVLTAAHCEFFV 365
QY	246	RKHTDVFNVKWVRAGSDKLGGFPSLA-VAKIIIIBFNPMY----PKNDIALMKLOFPITFS 301
DB	366	TREKVLBGWKYYAGTSNLHLQHPAAASTAEIIII---NSNYTDEEDDYDIALMRLSKPLTLS 422
QY	302	GTVRPICLPFEDELTATPLMTWGFTKQNGKGMSDILLQASVOVIDSTRCNADDAYO 361
DB	423	AHIHPACLPHMGQTFSLNETCWITFGFKTRTDTKTSPFFREVQVNLFIDFKKCNDYLVD 482
QY	362	GEVTEKMWACAGIPGGVDTCOGDSGGPLM-YQSQWHVVGVSVSGYGCGGPSTPGVVTKV 420
DB	483	SYLTPRMWCAGDLHGGRSDCOGDSGGGPLVCENRWIAGVTSNGTGCGQRKEGVTKV 542
QY	421	SAYLNWIYN 429
DB	543	TEVLPFIWS 551

```

RESULT 4
ID Q812A6 PRELIMINARY; PRT; 453 AA.
ID Q812A6
AC Q812A6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transmembrane proteinase tmpr33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
[1]
RN SEQUENCE FROM N.A.
RA Rao N.V., Rao G.N., Hoidal J.R.;
RT "Genomic Organization of Murine Transmembrane Proteinases.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF479687; AAC033581.1; -

```

```

DR GO: GO:0016020; C-membrane; IEA.
DR GO: GO:0004363; F-1-chymotrypsin activity; IEA.
DR GO: GO:0005044; F-scarvenger receptor activity; IEA.
DR GO: GO:0004295; F-trypsin activity; IEA.
DR GO: GO:0006508; P-proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR002172; LDL receptor_A.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR InterPro: IPR001190; Srcr receptor.
DR Pfam: PF00057; ldl recept a; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR SQ SEQUENCE 453 AA; 49505 MW; 1EE7ECDCB3DD894 CRC64;

Query Match 29.4%; Score 688; DB 11; Length 453;
Best Local Similarity 37.2%; Pred. No. 5.8e-58;
Matches 162; Conservative 71; Mismatches 153; Indels 50; Gaps 14;

Qy 22 IMETRKVGIPIIIALLSIASIIIVVLIKVLDD---KYFLCGQPLHFIIPKQLCDGE 78
Db 44 LPLKFFPIIIVIGIIATLALA-----IGLGHFDCSGKY--RCHSFKCIELTARCDGV 95
Qy 79 LDCPLGDEBHCVKSPPEGPAVARLSKDRSTLQVLDSATGNWFSACDFNTEALAEATAC 138
Db 96 SDCKNAEDVRC-----VRVSGQRAALQVFTAAA--WRTWCSSDDWKSHYAKIAC 142
Qy 139 RQMGYGS-----KPTTRAVEIGPDQLDVVEITENSQELRMNRSSGPECLSGSL 186
Db 143 AOLGFPYSVSSDHLRDALBEQFGQDFVSNHLLSDDKVTAHHSVVMREG---CTSGHV 199
Qy 187 VSLHCLACG-KSLKTRRVVCGEASVDSWPKVYSIQYDKOHVCGSGTLDPHVWLTAARCF 245
Db 200 VTLLKCSAGTRTGYSPIRVGNWSSLTQFWQVSLQFGYVHLCGSGSIITPLWIVTAAHCV 259
Qy 246 RKHTDVFNVKVRAGSKL--GSPFSLAVAKIIIEFNPMY-PK--DNDIALMKIQFELTF 300
Db 260 YDLVHPKSVTVQGLVSLMDSPVPSHLVEXII---YHSKYKPKRLGNDIALMKLSEPLTF 316
Qy 301 SGTVRPICLPFPOBELTTPATPLMIIGWGFTRKQNGKMSDILLQASVOVIVSTRCNADDAY 360
Db 317 DETIQIPLNSBENFPDGLKCLWTSNGAT-EDGDGASPVLNHAAPVLIISNKTICNHRDVI 375
Qy 361 QGVETKMCAGIPEGGVDPTRCQSDSGPLMYQSDQ-WHVUVGIVSWVCGCGGPGSTPGVYTK 419
Db 376 GGLISPSMLCAGYLKGGVDSQCGSDSGPLVQQRERLKLKLVGATSPGIGCAEVNKPQVYTR 435
Qy 420 VSAYLNWYVNVKAEI 435
Db 436 ITSFLDWIHEQLERDL 451

```

RESULT 5	
Q9BYEL	
ID	PRELIMINARY; PRT; 537 AA.
AC	Q9BYE1
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	Mosaic serine protease.
GN	MSPS.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21167393; PubMed=11267681;
 RA Kim D.R., Sharnin S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 RT without a transmembrane domain from human lung.";
 RL Biochim. Biophys. Acta 1518:204-209 (2001).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB048797; BAB39742.1; -;
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.087; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srrc_receptor.
 DR Pfam; PF00057; ldl_receptor; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0287; SRR2; 1.
 DR PROSITE; PS0240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match 29.3%; Score 686.5; DB 4; Length 537;

Best Local Similarity 38.6%; Pred. No. 18-57; Matches 142; Conservative 56; Mismatches 147; Indels 23; Gaps 7;

QY 75 CDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSDATGNWFSACDNPTEALA 134
 DB 181 CDGVVDCVCKLSDELGC-----VRFWDKSLKLYSGSSHQWLPICSSNNWDSYS 229
 QY 135 ETACQMGVSSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGCPCLSGSLVSLHCLAC 194
 DB 230 EKTCTQLGPES--AHRTEVAHRDFANSFSLRYNSTIQESLHRSHCPQRSYISLQCSHC 287
 QY 195 GSKSLKTPRVVGGESASVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCF--RKHTDVF 252
 DB 288 GURMTGRIVGGALASDSKWPQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRKYLE 347
 QY 253 NNKVRAGSKLGSFPFSLA-VAKIIIEFNPMY---PKNDIALMKLOPLTFSGTVRPI 308
 DB 348 GKWVAGTSLNHLQPEAASIAEII---NSNYTDEDDYDIALMRLSKPLTISAHPAC 404
 QY 309 LPFFDEELTPATPLWIGWFTKQNGKXMSDILLQASVQVIDSTFCNADDAVQGEVTEKM 368
 DB 405 LPMHGQTFSLNCTWITGFKTRETDDKTSPLREVQVNLIDFKCNLYLYDSVLTFRM 464
 QY 369 MCAGIPEGGVDTCCGDSGGPLM-YQSDQHVGVIVSWGCGGSPSTPGVYTKVSAYLNI 427
 DB 465 MCAGDLHGGRDSCQDSGGPLVCEQNNRWYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWI 524
 QY 428 YNVWAEEL 435
 DB 525 YSRMESEV 532

RESULT 6

Q86YM4 PRELIMINARY; PRT; 558 AA.

AC Q86YM4;
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Transmembrane protease serine 6.
 GN TMPSR56.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Park T.J., Park W.J.
 RT "Homo sapiens transmembrane protease, serine 6 (TMPSR56) mRNA.";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY190317; AAC38062.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srrc_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0287; SRR2; 1.
 DR PROSITE; PS0240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Protease.
 SQ SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

Query Match 29.3%; Score 686.5; DB 4; Length 558;

Best Local Similarity 35.1%; Pred. No. 11e-57; Matches 149; Conservative 74; Mismatches 163; Indels 39; Gaps 11;

QY 22 IPMETPR--KVGIPITIIALLSLASIIIVVVLKIVLDKYFLCGQP-LHFIPKQLC--- 75
 DB 149 LPRFTWREGQKQLFLGCVLLIALVLSLIL-----PQFWQGYTGIRYKEQRESCEPH 202
 QY 76 ----DGEELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSDATGNWFSACDNPTE 131
 DB 203 AVRDDGVDCVCKLSDELGC-----VRFWDKSLKLYSGSSHQWLPICSSNNWMD 251
 QY 132 ALAETACRMGVSSKPTFRAVEIGPDQDLVVEITENSQELMRNNSGCPCLSGSLVSLHC 191
 DB 252 SYSEKTCQLGPES--AHRTEVAHRDFANSFSLRYNSTIQESLHRSECPQRSYISLQC 309
 QY 192 LACGKSLKTPRVVGGESASVDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCF--RKHT 249
 DB 310 SHCGLRAMTGRIVGGALASDSKWPQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRK 369
 QY 250 DVFNWVRAGSKLGSFPFSLA-VAKIIIEFNPMY---PKNDIALMKLOPLTFSGTVR 305
 DB 370 VLEGWKVAGTSLNHLQPEAASIAEII---NSNYTDEDDYDIALMRLSKPLTISAHT 426
 QY 306 PICLPFDEELTPATPLWIGWFTKQNGKXMSDILLQASVQVIDSTFCNADDAVQGEVTE 365
 DB 427 PACLPMHGQTFSLNCTWITGFKTRETDDKTSPLREVQVNLIDFKCNLYLYDSVLT 486
 QY 366 EKMKCAGIPEGGVDTCCGDSGGPLM-YQSDQHVGVIVSWGCGGSPSTPGVYTKVSAYL 424
 DB 487 PRMCAGDLHGGRDSCQDSGGPLVCEQNNRWYLAGVTSWGTGCGQRNKPQVYTKVTEVL 546
 QY 425 NWIYN 429
 DB 547 PWIYS 551

RESULT 7

Q9DGR2 PRELIMINARY; PRT; 767 AA.
 ID Q9DGR2

```
AC Q9DGR2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Embryonic serine protease-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis."
RL Gene 252:209-216(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSP; P00766; ICHG.
DR MEROPS; S01.043; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001454; Peptidase_S1.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 8.
DR SMART; SM00020; tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 8.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE596E CRC64;

Query Match 29.3%; Score 686.5; DB 13; Length 767;
Best Local Similarity 38.3%; Pred. No. 1.7e-57;
Matches 145; Conservative 58; Mismatches 153; Indels 23; Gaps 7;

QY 62 CQOPLHPIPRKQLDGLDCEHCHVKSPPGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 395 CGSSVSCVLSQWCDGVSDCPYGEDEMSCVSLYPAD-----PQLQVYSTSVSAW 443
QY 122 FSACPDNFTALAEATACRMGYSSKPTFRAVEI-----GPDQDLVVEITENSQELMRNS 177
DB 444 LPVCSQYNDPGRACQDFGNGSSYNYRDTLMSPYAPNGYFKLYSGYWSKFTYVOY 503
QY 178 SGPCLSGSLVSLHCLACGKSLKT--PRVVGGEASVDSWPQVSIQYDKQHVCGGSLDP 235
DB 504 SSYCVSGNVSLHCLISCGVSNLSRVISVGGTFANLGNWPQVNLQYITGVLCGGSIIP 563
QY 236 HWLVTAAHC-FRKHTDVENVKVRAGSKLGSPPSLA---VAKIIIEFNPMYPKNDIAL 291
DB 564 KWIIVTAHCVGSYSASGWRVAGTLTKPSYKNSAYFVERIIVHPGKYKTYDNDIAL 623
QY 292 MKLQPLTFSGTVRPICLPFFDEELTPATPLWIGFTKQNGKMSDILLQASVQVIDS 351
DB 624 MKLRDEITFGYTTQVCLPNSGMFEAGTTTWISGMGSTYB-GGSVSTYLQYAAIPLIDS 682
QY 352 TRCNADDAQGEVTEKMKAGIPGCVDTCCGDSGGLM-YQSQWHVGVISVGYCGGG 410
DB 683 NVNCGSYVNGITSSMTCAGVLSGGVDTCCGDSGGLNKGNGFWLVLVGTSGDGOAR 742
QY 411 PSTPGVYTKVSAYLNWIYN 429
```

```
Db 743 ANKPGVYGVNTTFLEWIS 761
RESULT 8
Q96T73 PRELIMINARY; PRT; 492 AA.
ID Q96T73
AC Q96T73;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epitheliasin.
GN TMRSS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223025; PubMed=11328890;
RA Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
RA Hoidal J.R.;
RT "Cloning and characterization of the cDNA and gene for human
epitheliasin."
RL Eur. J. Biochem. 268:2687-2699(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AF329454; AAK53559.1; -.
DR HSP; P00761; IAN1.
DR GO; GO:0015020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match 28.6%; Score 670.5; DB 4; Length 492;
Best Local Similarity 38.8%; Pred. No. 3.3e-56;
Matches 146; Conservative 59; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCEPLGEDEHCHVKSPPGPAVAVRLSKDRSTLQVLDSATGNWFSACPDNFTAL 134
DB 133 CDGVSHCPGGEDEHCHVKSPPGPAVAVRLSKDRSTLQVLDSATGNWFSACPDNFTAL 181
QY 135 ETACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSG-----PCL 182
DB 182 RAACKDMGY--KNFYSSQ-----GIVDSSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSVSLHCLACGKSL---KTPRWVGGEASVDSWPQVSIQYDKQHVCGGSLDPHWVL 239
DB 233 SKAVVSLRCLACGVLNLSRQSLVGGESALPCAWPQVSLHVQNVHVCVGSIIITEWIV 292
QY 240 TAAHCFRKH-TDVFNKKVRAGSKLGSF---PSLAVAKIIIEFNPMY---PKNDIAL 291
DB 293 TAAHCFVEKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
```

QY 292 MKQLPFLPSTGTRVTRPCLPFFDELTATPWLIGWFTKONGKMSDILLOASVQVIDS 351
DB 349 MKQLPFLPSTGTRVTRPCLPFFDELTATPWLIGWFTKONGKMSDILLOASVQVIDS 407
QY 352 TRCNADDAVQGVTEKMKAGIPGEGVDTCCGDSGGLMYQSDQ-WHVVGIVSNGYCGG 410
DB 408 QRCNRYVYDNLITPAMICAGFLQGVNDSQCGSGLPVTSKNNIWLIGTWSGSGCAK 467
QY 411 PSTPGVYTKVSAYLNNWVWKA 434
DB 468 AYRPGVYGVNVTFTDWIYRQWAD 491
RESULT 9
Q7TN04 PRELIMINARY; PRT; 490 AA.
AC Q7TN04;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Tmrs22 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano P.J., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan N.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Schmutz J., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054348; AAHS4348.1; -. 54650B028417665A CRC64;
SQ SEQUENCE 490 AA; 53525 MW; 54650B028417665A CRC64;

Query Match 27.7%; Score 648; DB 11; Length 490;
Best Local Similarity 37.1%; Pred. No. 5e-54;
Matches 155; Conservative 61; Mismatches 162; Indels 40; Gaps 14;
QY 39 LSIASIIIVVILKVIDKY-----FLGQPLHPIPRKQLCDGELDPLGDEEHC 90
DB 88 LALGTVLTGAAYVALLVFWDSNCTSMGCGSGTCLSSSLWCDGVAHPCNGEDNRC 147
QY 91 VKSPFGPAVAVRLSKDRSTLQVLDSATCNWFSACFDNFTEALAEATACRQMSYKPTFR 150
DB 148 -----VRLYQGSFILQVSSQKAWYPVQDDWSESYGRAACKDMGY--KNNFY 194
QY 151 AVEIGDQD--LDVVEITNSOELMRNS---SGCLSGSLVSLHCLAG--KSLK--TPRV 203
DB 195 SSQGIPOQSGATSFMKLVNSSGVDLYKKLYHSDSCSRMVMVSLRCIEGVRSVKQSGRI 254

QY 204 VGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTAACFRKH-TDVFNWVKVAG--S 260
DB 255 VGLNASPGDWPQVSLHVQGVHVGCGSIITPEWIVTAACVCEELPSSPRYTAAGILR 314
QY 261 DKLGSPSL-AVAKIIIEFNPMY---PKNDIALMKLOPLTFSTGTVRPICLPFFDEEL 316
DB 315 QSLMFVGSRSHVEKVI---SHPNYDSKTKNDIALMKLOPLTFSTGTVRPICLPFFDEEL 371
QY 317 TPATPWLIGWFTKONGKMSDILLOASVQVIDSTRCNADDAVQGVTEKMKAGIPG 376
DB 372 DLQECWISGWGATYK-GKTSVLNAAVPLIEPSKCNKYINNLIITPAMICAGFLQ 430
QY 377 GVTCCGDSGGLM-YQSDQWVVGIVSNGYCGGSPSTGVTYKVSAYLNNWVWKA 433
DB 431 SVDSQCGDSGGLVTLKNGIWLIGTWSGSGCAKALRFGVYGVNVTFTDWIYQWRA 488
RESULT 10
Q920K3 PRELIMINARY; PRT; 490 AA.
AC Q920K3;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Tmrs22.
GN Tmrs22.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RT "Tmrs22, Rat."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB073550; BAB70683.1; -.
DR HSPB; P00761; IAN1.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0004263; F.chymotrypsin activity; IEA.
DR GO; GO:0008233; F.peptidase activity; IEA.
DR GO; GO:0005044; F.scavenger receptor activity; IEA.
DR GO; GO:0004295; F.trypsin activity; IEA.
DR GO; GO:0006508; P.proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00068; LDLA_2; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 490 AA; 53515 MW; 2BC691551CAC409A CRC64;

Query Match 27.6%; Score 646; DB 11; Length 490;
Best Local Similarity 38.8%; Pred. No. 7.9e-54;
Matches 151; Conservative 54; Mismatches 148; Indels 36; Gaps 13;
QY 62 CGOPLHPIPRKQLCDGELDPLGDEEHCVKSPFGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 119 CGSSGTCLSSSLWCDGVAQCPNGKDNRC-----VRLYGTSTLQVYSSQKAW 167
QY 122 FSACFDNFTEALAEATACRQMSYKPTFRFAVEIGDQ-----DLVVEITNSOEL-RMRN 176
DB 168 YPVQDDWSESYGRAACKDMGY--KNSFYSSQGIPOQSGATSFMKLVNSSGVDLYKKLY 225

```
QY 177 SSGPCLSGLSLVSLHCLACG-KSL-KTPRVVGGEEASVDSWPNQVSIQYDKQHVCGGSLD 234
Db 226 HSDSCSRVWSLRICBGRVSRQSRIVGSGTASPDGWPQVSLHVQGIHVCGGSLIT 285
QY 235 PHWLTAARHCRKH-TDVFNWVKVRAGSDK-----LGSPFSLAVAKIIIEFNPMYPKN---DIALMKLQFP 297
Db 286 PEWIVTARHCVBEPLSSPRYWTAFAGILKQSLMFYGS--RHQVEKVI---SHPNVDSKTK 340
QY 286 DNDIALMKLQPLTFSQTVRPICLPFFDEBLTPATPLMIIGWFTKQNGKMSDILLQASVOVIDSTRCNAD 345
Db 341 NNDIALMKLQPLTAYDNDVVKVCLPNPGWMLDLAQECWISGWGATYEK-GRTSDVLNAAM 399
QY 346 VOVIDSTRCNADADYQGEVTEKMCAGIPEGVDTCQDSDGGPLM-YOSDQWVGVISW 404
Db 400 VPLIEPSKNSKYLYNNLITAMICAGFLQSDVSCQSDSGGLVTLKNEIWLWIGTISW 459
QY 405 GYGCGGSPSTPGYTKVSAYLNIWYNWKA 433
Db 460 GSGCAKAYRPGVGNVFTDWTIYQOMRA 488

RESULT 11
Q8CU17 Q8CU17 PRELIMINARY; PRT; 445 AA.
AC Q8CU17;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Adrenal mitochondrial protease long variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEBH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537098; AA06757.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001354; Peptidase_S1.
DR InterPro; IPR001190; Srrc_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 445 AA; 48440 MW; BE3F56D8372ED988 CRC64;

Query Match 27.2%; Score 636.5; DB 11; Length 445;
Best Local Similarity 36.2%; Pred. No. 5.8e-53;
Matches 135; Conservative 62; Mismatches 141; Indels 35; Gaps 10;

QY 79 LDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTEALAEATAC 138
Db 81 LSCPGVSSEKLLPSLPK-AVSRFNGEDLLEVVQVREARPDLLVCHEGNPAIAGWHIC 138
QY 139 RQMGYSKPTFRAVEIGDPQDLVDVEITENSQELMRNS-----SGPCLSGSL 186
Db 139 QSLGFRLTQKAVNL-----SDIKL-----NRSQEFALQSARPGSLVBEAWQFSTNCPGRI 191
```

```
QY 187 VSLHCLACGKSLXTPRVVGGEEASVDSWPNQVSIQYDKQHVCGGSLDPHWLTAAHC-- 244
Db 192 VSLKCEGCGARPASRIVGGQAVASGRWPMQASWMLGSRHTCGSVLAPYVVVTAACMY 251
QY 245 -FRKHDTDFVFNWVKVRAG-----SDKLGSPFSLAVAKIIIEFNPMYPKN---DIALMKLQFP 297
Db 252 SFRL-SRLSSWRVHAGLVSHSAVRQHGTVWEKLIIP---HPLYSAQNHVDYDVALQLRTP 307
QY 298 LTSSTGVTRPICLPFFDEBLTPATPLMIIGWFTKQNGKMSDILLQASVOVIDSTRCNAD 357
Db 308 INFSDTIVSAYCLPAKEQHFPGSQCVWVGWHTDPSHTSHSSDTLQDTMVPFLSLDLCNSS 367
QY 358 DAYQGEVTEKMCAGIPEGVDTCQDSDGGPLM-YOSDQWVGVISW 404
Db 368 CMTSGALTHMLCAGYLDGRADACQSDSGGLVCPGSDTWHLVGVSWGRCGAEPRPGV 427
QY 417 YTKVSAYLNIWYN 429
Db 428 YAKVAEFLDWIHD 440

RESULT 12
Q8CDRO Q8CDRO PRELIMINARY; PRT; 455 AA.
AC Q8CDRO;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Transmembrane protease.
GN TMPSRSS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK029714; BAC26577.1; -.
DR MGD; MGI:1933407; Tmprss5.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srrc_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 455 AA; 49669 MW; BE22EB2E7503C74B CRC64;

Query Match 27.0%; Score 631.5; DB 11; Length 455;
Best Local Similarity 36.2%; Pred. No. 1.8e-52;
Matches 135; Conservative 63; Mismatches 140; Indels 35; Gaps 10;

QY 79 LDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTEALAEATAC 138
Db 91 LNCPGVSREELDPSLPK--TVSRFNGEDLLEVVQVREARPDLLVCHEGNPAIAGWHIC 148
QY 139 RQMGYSKPTFRAVEIGDPQDLVDVEITENSQELMRNS-----SGPCLSGSL 186
```

```

Db      149 KSLGHLRLTQHKAVNL---SDIKL---NRSQEPAQLSARPGGLVEESWKSANCPGSGRI 201
Qy      187 VSLHCLACGKSLKTRPVVGGEEASVDSPWQVSIQYDKQHVCGGSIILDPHVLTAAC-- 244
Db      202 VSLKCKSECGARPLASRIVGGQAVASGRVPWQASVNLGSRHTCGASVLAPHVWVTAACHMY 261
Qy      245 -PRKTDVFNWVRAGSKLGS---FPSLAVAKIIIEFNPMYKDN---DIALMKLOFP 297
Db      262 SFL--SRSSWVHAGLVSHGAVRQHQGTWVEKIIP---HPLYSAQNHDYDVALQLRTP 317
Qy      298 LTFSTGVTRPCLPFFDEBLTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCAD 357
Db      318 INFSDTVGAVCLPAKEQHPQSGCWVSGWGHDTFSTHSSDTLQDTWVPLLSLYLNCSS 377
Qy      358 DAYQGEVTEKMKCAGIPEGVDTCQDGGPLMYQS-DQWVHVGVSVWGYCGGPGSTPGV 416
Db      378 CMYSALTRMLCAGYLDGRADACQDGGPLVCPSGDTWHLVGVVSWGRGCABPNPGV 437
Qy      417 YTKVSAYLNIWYN 429
Db      438 YAKVAEFLDWIHD 450

RESULT 13
Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF537099; AN06758.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase SLA.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; TRYPSIN; 1.
DR PROSITE; PS00287; SRCR 2; 1.
DR PROSITE; PS0240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

```

```

Query Match      26.1%; Score 612; DB 11; Length 371;
Best Local Similarity 34.7%; Pred. No. 1.1e-50;
Matches 131; Conservative 62; Mismatches 138; Indels 46; Gaps 11;

Qy      87 EHCVKSPFEGPA-----VAVLSKDRSTLQVLSATGNWFSACDFNTEALA 134
Db      2 EIRCTEEG-GPGIFRAELGDDQQPSFRINGEDLLLEQVRAEPLDMLLVCHEGWNPALG 60
Qy      135 ETACRMGYSSKFTFAVEIGPDQDLDVVEITENSQELRMNS-----SGPCL 182

```

```

Db      61 MHICQSLGFRLTQHKAVNL---SDIKL---NRSQEPAQLSARPGSLVEEAWQSTNCP 113
Qy      183 SGLVSLHCLACGKSLKTRPVVGGEEASVDSPWQVSIQYDKQHVCGGSIILDPHVLTAAC 242
Db      114 SGRIVSLKSECGARPLASRIVGGQAVASGRVPWQASVNLGSRHTCGGSLVAFYVVTAA 173
Qy      243 HC---PRKHTDVFNWVRAG---SDKLGFPSLAVAKIIIEFNPMYKDN---DIALMK 293
Db      174 HCMYSFL--SRSSWVHAGLVSHGAVRQHQGTWVEKIIP---HPLYSAQNHDYDVALQL 229
Qy      294 LQPLPFSGTVPICLPFFDEBLTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTR 353
Db      230 LRTPIINFSDTVGAVCLPAKEQHPQSGCWVSGWGHDTFSTHSSDTLQDTWVPLLSIDL 289
Qy      354 CNADDAYQGEVTEKMKCAGIPEGVDTCQDGGPLMYQS-DQWVHVGVSVWGYCGGPGS 412
Db      290 CNSSCMYSALTRMLCAGYLDGRADACQDGGPLVCPSGDTWHLVGVVSWGRGCABPN 349
Qy      413 TPGVYTKVSAYLNIWYN 429
Db      350 RGVYAKVAEFLDWIHD 366

RESULT 14
Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protease (Fragment).
GN PRSS7 OR A130037D21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK038356; BAC29973.1;
DR MGD; MGI:1197523; Pssr7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000839; CUB; 1.
DR InterPro; IPR000903; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase SLA.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; tryptsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; LDua; 1.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.

```

DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 1.
 DR PROSITE; PS00740; MAX_1; 1.
 DR PROSITE; PS00060; MAX_2; 1.
 DR PROSITE; PS02087; SECR_2; 1.
 DR PROSITE; PS02040; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 FT NON TER 1
 SQ SEQUENCE 777 AA; 87314 MW; A18E2F4EC06D3A8 CRC64;

Query Match 23.3%; Score 545; DB 11; Length 777;
 Best Local Similarity 32.5%; Pred. No. 9.8e-44;
 Matches 124; Conservative 72; Mismatches 148; Indels 38; Gaps 13;

QY 69 IPRKQCDGELDCPLGDEEHCVKSFEGPAVAVRLSKDSTLQVLD-SATGNWFSACFD 127
 Db 410 IPLGNLDCSYPHCRDGSDEASCVRP-----LNGTRSNGLVQFNHSHIACAE 459
 QY 128 NFEALAEATACRMQYSSKPTFRAVEI---GPDQDLVVEITENSQELRNSSGPGCLSG 184
 Db 460 NWTTCISNEVCHLLGLGSANSSMPSITGGGP-----FVRVYQAPNGSLIILTPSLQCSQD 514
 QY 185 SLVSLHC--LACGSLKT-----PRVVGEEASVDSWPQVSIQYDKH-----VCGGSILD 234
 Db 515 SLILLQCNHCKGCKKVTQKVSPIVGGSDAQAGWPVWVYHRDRSTDRLLCGASLVS 574
 QY 235 PHWVLTAAHC-FRKHDTVFNNKVRAGSKLGSFPFLAVAKIII--IBFNMPYP---KDND 288
 Db 575 SDWLVSAAHCYVRNLDPTFTAVLGLHMQSNLTSPOVRRVVDQIVINPHYDERRKVD 634
 QY 289 IALMKLQPLFTSGTVRPICLPFFDEELTPATPLWIGWFTKONGKMSDILLOASVQV 348
 Db 635 IAMEHLEPKVNYTDYIOPICLPEENQIFIPGRTCSAGWGYDKINGSTV-DVLKEADVPL 693
 QY 349 IDSTRCNADDAVQGEVEKMKCAGIPGGVDVTCQDGGPLMYQ--SDQHVVGIVSWGYG 407
 Db 694 ISNEKCO-QQPEYNITTESMTICAGYEGGIDSCQDGGPLMCGENNRWFLVGVTSFGVQ 752
 QY 408 CGGPSTPGVYTKVSAYLNWIYN 429
 Db 753 CALPNHGGVYVRVSOQFIEWIHS 774

RESULT 15

Q8VHJ4 ID Q8VHJ4 PRELIMINARY; PRT; 417 AA.
 AC Q8VHJ4;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Airway trypsin-like protease.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_taxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Trachea;
 RA Hansen I.A.; Fasnacht M.; Hammer F.; Schammann M.; Allolio B.;
 RT "Cloning and characterization of RAT, the airway trypsin-like protease
 of Rattus norvegicus."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AF453776; AAL50817.1; -.
 DR HSP; P00761; IANI.
 DR MEROPS; S01.047; -.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase S1A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0722; CHYMOTRYPSIN.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS00024; SEA; 1.
 DR PROSITE; PS00240; TRYPsin DOM; 1.
 DR PROSITE; PS00134; TRYPsin HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 417 AA; 46287 MW; DB9504158B018E21 CRC64;

Query Match 22.9%; Score 535; DB 11; Length 417;
 Best Local Similarity 34.4%; Pred. No. 3.1e-43;
 Matches 129; Conservative 48; Mismatches 98; Indels 100; Gaps 14;

QY 89 HCYKSPFEGPAVAV-----RLSKD-----RSTLQVLDSATGNWFSACFDNFEA 132
 Db 101 HVVKLRKEGSGVADVVMKFRSSKRNKKAIKTRIQSVLQRL--SSSGN----- 147
 QY 133 LAETACRMQYSSKPTFRAVEIGP-----DQDLVVEITEN--SQELRNSSGPGCL 182
 Db 148 -----LEIAPNGITSLTDQD-----TENVLTOE----- 171
 QY 183 SGLVSLHCLACGK-----SLKTPRVVGEEASVDSWPQVSIQYDKHVCVGSILDPHW 237
 Db 172 -----CGARPDLLTSEERIIGCTQAEIGDWPQVSLQNNVHHCGLTSLNLW 220
 QY 238 VLTAAHCFKHTDVFNNKVRAGSKLGSFPFLAVAKIII---EFNMYPKNDIALMKL 294
 Db 221 VLTAAHCFHSYSPNQOWTATFGVSTIS--PRLRVRVRAILAHAEYNSI--TRNDIAVQL 277
 QY 295 QPPLTFSGTVRPICLPFFDEELTPATPLWIGWFTKONGKMSDILLOASVQVIDSTRC 354
 Db 278 DRVTTTRIHRVCLPAAQNIMPDSVAVYVGWSLTYGNTVTN--LQGEVRIIVSSEVC 336
 QY 355 NADDAVQGEVTERKMKCAGIPGGVDVTCQDGGPLMYQSDQ--WHVVGIVSWGYCGGSPS 412
 Db 337 NEPAGYGGSVLPGLMCLAGVRSAGVADACQDGGPLVQEDTRRLWFVVGIVSWGYCQGLPN 396
 QY 413 TPGVYTKVSAYLNWI 427
 Db 397 KPGVYTRVTAAYRNWI 411

Search completed: June 7, 2004, 08:42:05

Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:41:12 ; Search time 48 Seconds
(without alignments)
2549.632 Million cell updates/sec

Title: US-10-030-688-2

Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVPLRKP.....VYTRVSAYLNWYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/FCRUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2338	99.8	435	12	US-09-968-415-6
2	2338	99.8	435	14	US-10-180-719-6
3	2337	99.8	437	9	US-09-851-588-8
4	2337	99.8	437	15	US-10-295-027-779
5	2337	99.8	437	15	US-10-295-027-791
6	2337	99.8	437	15	US-10-295-027-831
7	2337	99.8	437	15	US-10-295-027-1196
8	2337	99.8	437	15	US-10-173-989-89
9	2319	99.0	437	15	US-09-776-191-4
10	2319	99.0	437	15	US-10-156-214A-4
11	2297.5	98.1	432	10	US-09-888-257A-7
12	2297.5	98.1	432	10	US-09-946-374-275
13	2297.5	98.1	432	12	US-10-206-915-330
14	2297.5	98.1	432	12	US-10-199-670-330
15	2297.5	98.1	432	12	US-10-201-858-330

16 2297.5 98.1 432 12 US-10-205-890-330
17 2297.5 98.1 432 12 US-10-208-024-330
18 2297.5 98.1 432 12 US-10-201-853-330
19 2297.5 98.1 432 12 US-10-063-745-112
20 2297.5 98.1 432 12 US-10-063-512-112
21 2297.5 98.1 432 12 US-10-063-513-112
22 2297.5 98.1 432 12 US-10-063-515-112
23 2297.5 98.1 432 12 US-10-063-549-112
24 2297.5 98.1 432 12 US-10-063-549-112
25 2297.5 98.1 432 12 US-10-063-569-112
26 2297.5 98.1 432 12 US-10-174-581-330
27 2297.5 98.1 432 12 US-10-176-483-330
28 2297.5 98.1 432 12 US-10-176-749-330
29 2297.5 98.1 432 12 US-10-176-914-330
30 2297.5 98.1 432 12 US-10-176-915-330
31 2297.5 98.1 432 12 US-10-006-485A-275
32 2297.5 98.1 432 12 US-10-013-907A-275
33 2297.5 98.1 432 12 US-10-015-499A-275
34 2297.5 98.1 432 12 US-10-063-553-112
35 2297.5 98.1 432 12 US-10-063-563-112
36 2297.5 98.1 432 12 US-10-063-594-112
37 2297.5 98.1 432 12 US-10-063-553-112
38 2297.5 98.1 432 12 US-10-063-554-112
39 2297.5 98.1 432 12 US-10-176-484-330
40 2297.5 98.1 432 12 US-10-180-550-330
41 2297.5 98.1 432 12 US-10-183-014-330
42 2297.5 98.1 432 12 US-10-187-738-330
43 2297.5 98.1 432 12 US-10-187-740-330
44 2297.5 98.1 432 12 US-10-187-883-330
45 2297.5 98.1 432 12 US-10-194-363-330

ALIGNMENTS

RESULT 1
US-09-968-415-6
; Sequence 6, Application US/09968415
; Publication No. US20020086334A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/968,415
; FILING DATE: 26-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/659,151
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555

```

;
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 435 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: COLNNOT13
;   CLONE: 1337018
;   SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-968-415-6

Query Match          99.8%; Score 2338; DB 12; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-226;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYIF 60
DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYIF 60
QY 61 LCQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLD SATGN 120
DB 61 LCQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNTEALAEACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELMRNSSGP 180
DB 121 WFSACFDNTEALAEACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
QY 241 AAHCFRKHDTVFNKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
DB 241 AAHCFRKHDTVFNKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVCYGGGPGTGYTKV 420
DB 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVCYGGGPGTGYTKV 420

```

RESULT 2

```

US-10-180-719-6
; Sequence 6, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

```

;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/180,719
;   FILING DATE: 25-Jun-2002
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/09/008,271
;     FILING DATE: 16-Jan-1998
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mohan-Peterson, Sheela
;   REGISTRATION NUMBER: 41,201
;   REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650-855-0555
;   TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 435 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: COLNNOT13
;   CLONE: 1337018
;   SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-10-180-719-6

Query Match          99.8%; Score 2338; DB 14; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-226;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYIF 60
DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYIF 60
QY 61 LCQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLD SATGN 120
DB 61 LCQPLHFIPIRKLCDGELDCPLGEDEEHCVKSFPEGPAPAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNTEALAEACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELMRNSSGP 180
DB 121 WFSACFDNTEALAEACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
QY 241 AAHCFRKHDTVFNKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
DB 241 AAHCFRKHDTVFNKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVCYGGGPGTGYTKV 420
DB 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVCYGGGPGTGYTKV 420
QY 421 SAYLNWIYNNWKAEL 435
DB 421 SAYLNWIYNNWKAEL 435

```

RESULT 3

```

US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.

```

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
 ; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
 ; FILE REFERENCE: A-68829-1/DJB/JUD/AWS
 ; CURRENT APPLICATION NUMBER: US/09/851,588
 ; CURRENT FILING DATE: 2001-09-24
 ; PRIOR APPLICATION NUMBER: US/09/642,252
 ; PRIOR FILING DATE: 2000-08-17
 ; PRIOR APPLICATION NUMBER: US/09/656,002
 ; PRIOR FILING DATE: 2000-09-06
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-851-588-8

Query Match 99.8%; Score 2337; DB 9; Length 437;
 Best Local Similarity 100.0%; Pred. No. 1.6e-226;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLLIKVILDKYFL 61
 Db 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLLIKVILDKYFL 63
 QY 62 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 121
 Db 64 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNTEALAEATACRMQGYSSKPTTFRAVEIGDPDLDVVEITENSQELRMNSSGPC 181
 Db 124 FSACFDNTEALAEATACRMQGYSSKPTTFRAVEIGDPDLDVVEITENSQELRMNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHDTVFNWKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
 Db 244 AHCFRKHDTVFNWKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
 QY 302 GTVRPICLPFFDEELTPATPLWIIWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAY 361
 Db 304 GTVRPICLPFFDEELTPATPLWIIWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAY 363
 QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 421
 Db 364 GEVTEKMCAGIPEGGVDTCCGDSGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 423
 QY 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 4
 US-10-295-027-779
 ; Sequence 779, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US/09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US/60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US/60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US/60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US/60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US/60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US/60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US/60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US/60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US/60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 779
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-295-027-779

Query Match 99.8%; Score 2337; DB 15; Length 437;
 Best Local Similarity 100.0%; Pred. No. 1.6e-226;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLLIKVILDKYFL 61
 Db 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIIALLSLASIIIVVLLIKVILDKYFL 63
 QY 62 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 121
 Db 64 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNTEALAEATACRMQGYSSKPTTFRAVEIGDPDLDVVEITENSQELRMNSSGPC 181
 Db 124 FSACFDNTEALAEATACRMQGYSSKPTTFRAVEIGDPDLDVVEITENSQELRMNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHDTVFNWKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
 Db 244 AHCFRKHDTVFNWKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
 QY 302 GTVRPICLPFFDEELTPATPLWIIWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAY 361
 Db 304 GTVRPICLPFFDEELTPATPLWIIWGFTKQNGKMSDILLOASVQVIDSTRCNADDAAY 363
 QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 421
 Db 364 GEVTEKMCAGIPEGGVDTCCGDSGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 423
 QY 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 5
 US-10-295-027-791
 ; Sequence 791, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel

364 GEVTEKMCAGIPGCGVDYTCQDSSGGLMYQSDQWHVVGVISWGYCGGSPSTGCVYTKUS 42

422 AYLNIWYNWKAEL 435
|||||

424 AYLNIWYNWKAEL 437

RESULT 6
US-10-295-027-831
; Sequence 831, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.

```

; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 831
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-295-027-831

      99.8%; Score 2337; DB 15; Length 437;
      Best Local Similarity 100.0%; Pred. No. 1-6e-226;
      Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      2  DFDSDQPLNSLDVKPLRKPRIPMETFRKVGIPPIIALSLASIIIVVLKIVLDKYFYL 61
Db      4  DFDSDQPLNSLDVKPLRKPRIPMETFRKVGIPPIIALSLASIIIVVLKIVLDKYFYL 63

QY      62  CGQPLHFTPRKQLCGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Db      64  CGQPLHFTPRKQLCGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123

QY      122  FSACFDNFTEALASTACRMQYSSKPTFRAVEIGPDDLDVVEITENSQELMRNNSGGPC 181
Db      124  FSACFDNFTEALASTACRMQYSSKPTFRAVEIGPDDLDVVEITENSQELMRNNSGGPC 183

QY      182  LSGSLVSLHCLACSKSLKTRVVGGEASVDSPWPQVSIQYDKQHVCGGSGSILDPHWLTA 241

```

Db 184 LGSLSVLSHCLACGSLKTPRVVGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
QY 242 AHCPRKHTDVFNKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCPRKHTDVFNKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLOASVQVIDSTRCNADDAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLOASVQVIDSTRCNADDAYQ 363
QY 362 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 421
Db 364 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 423
QY 422 AYLNWIYNNWKAEL 435
Db 424 AYLNWIYNNWKAEL 437

RESULT 7

US-10-295-027-1196
; Sequence 1196, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1196
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1196

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1,6e-226;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 63
QY 62 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Db 64 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPQDLDVVEITENSQELVRNMSGPC 181
Db 124 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPQDLDVVEITENSQELVRNMSGPC 183
QY 182 LGSLSVLSHCLACGSLKTPRVVGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 241
Db 184 LGSLSVLSHCLACGSLKTPRVVGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
QY 242 AHCPRKHTDVFNKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCPRKHTDVFNKVRAGSKLGSFPLSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLOASVQVIDSTRCNADDAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGGKMSDILLOASVQVIDSTRCNADDAYQ 363
QY 362 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 421
Db 364 GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 423
QY 422 AYLNWIYNNWKAEL 435
Db 424 AYLNWIYNNWKAEL 437

RESULT 8

US-10-173-999-89
; Sequence 89, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-89

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1,6e-226;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 63
QY 62 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121

Db 64 CGOPLHPIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 181
 Db 124 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
 Db 304 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 363
 QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGSPSTPGVYTKVS 421
 Db 364 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGSPSTPGVYTKVS 423
 QY 422 AYLNWIYNVWKAE 435
 Db 424 AYLNWIYNVWKAE 437

RESULT 9
 US-09-776-191-4
 ; Sequence 4, Application US/09776191
 ; Publication No. US20030119168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwin L. Madison
 ; APPLICANT: Edgar O. Ong
 ; APPLICANT: Jiunn-Chern Yeh
 ; APPLICANT: Corvas International, Inc.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
 ; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
 ; FILE REFERENCE: 24745-1607
 ; CURRENT APPLICATION NUMBER: US/09/776,191
 ; CURRENT FILING DATE: 2001-02-02
 ; PRIOR APPLICATION NUMBER: 60/213,124
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/234,840
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/179,982
 ; PRIOR FILING DATE: 2000-02-03
 ; PRIOR APPLICATION NUMBER: 60/183,542
 ; PRIOR FILING DATE: 2000-02-18
 ; PRIOR APPLICATION NUMBER: 09/657,968
 ; PRIOR FILING DATE: 2000-02-08
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-776-191-4

Query Match 99.0%; Score 2319; DB 10; Length 437;
 Best Local Similarity 99.5%; Pred. No. 1e-224;
 Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLKIVLDKYFL 61
 Db 4 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLKIVLDKYFL 63
 QY 62 CQOPLHPIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CQOPLHPIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 181

Db 124 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
 Db 304 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 363
 QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGSPSTPGVYTKVS 421
 Db 364 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGSPSTPGVYTKVS 423
 QY 422 AYLNWIYNVWKAE 435
 Db 424 AYLNWIYNVWKAE 437

RESULT 10
 US-10-156-214A-4
 ; Sequence 4, Application US/10156214A
 ; Publication No. US20040001801A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwin L. Madison
 ; APPLICANT: Joseph Edward Sempke
 ; APPLICANT: George P. Vlasuk
 ; APPLICANT: Scott Jeffrey Kemp
 ; APPLICANT: Mallareddy Komandla
 ; APPLICANT: Daniel Vanna Siev
 ; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
 ; TITLE OF INVENTION: Thereof
 ; FILE REFERENCE: 24745-1611
 ; CURRENT APPLICATION NUMBER: US/10/156,214A
 ; CURRENT FILING DATE: 2002-05-23
 ; NUMBER OF SEQ ID NOS: 611
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-156-214A-4
 Query Match 99.0%; Score 2319; DB 15; Length 437;
 Best Local Similarity 99.5%; Pred. No. 1e-224; 2; Indels 0; Gaps 0;
 Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLKIVLDKYFL 61
 Db 4 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLKIVLDKYFL 63
 QY 62 CQOPLHPIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CQOPLHPIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 181
 Db 124 FSACFNFTALAEACRQMGYSKPTFRFAVEIGPDQDLVDVEITENSQELMRNNSGFC 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
 QY 242 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTDFVNMKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361

Db	304	GVFRILCLPFDEELTPATPLWIIGWFTKQNGKMSDILLQASQVWIDSTRNADDAYQ	363
QY	362	GEVTEKMWKAGIPPEGVDTCQDGGGLMTQSDQHWVVGIVSWGYGCGGPSTPGVTTKVS	421
Db	364	GEVTEKMWKAGIPPEGVDTCQDGGGLMTQSDQHWVVGIVSWGYGCGGPSTPGVTTKVS	423
QY	422	AYLNWIYNVWKAEL	435
Db	424	AYLNWIYNVWKAEL	437

RESULT 11
US-09-888-257A-7
; Sequence 7, Application US/09888257A
; Publication NO. US20030060612A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Polakis, Paul
; APPLICANT: Smith, Victoria
; APPLICANT: Wood, William I.
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Xemin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TREATMENT OF TUMOR
; FILE REFERENCE: P5002R1

```

Query Match      98.1%; Score 2297.5; DB 10; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-22;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY  2  DPDSQPSLSDVVKPLAKPRIPMETFRKVGPIPIIALLSLAGIIIVVVLKIVLDKYVFL 61
Db  4  DPDSQPSLSDVVKPLAKPRIPMETFRKVGPIPIIALLSLAGIIIVVVLKIVLDKYVFL 63

```

Qy	62	CQCPHFHPRXQLCDGELDCPLGEDDEHCVKSPPEGPAPAVRLSKDORSTLQVLDSATGNW	121
Dd	64	CQCPHFHPRXQLCDGELDCPLGEDDEHCVKSPPEGPAPAVRLSKDORSTLQVLDSATGNW	123
Qy	122	FSACFDNTEALAEATACRMQYSKPTFRAVEIGPDODLDVWEITENSBELMRNSSGGC	181
Dd	124	FSACFDNTEALAEATACRMQYS-----RAVEIGPDODLDVWEITENSBELMRNSSGGC	178
Qy	182	LSSGLSVSLHCLACGKSUKTPRVVGEEASVDSPWPQVSIQYDXQHVCVGSILDPHHVLTA	241
Dd	179	LSSGLSVSLHCLACGKSUKTPRVVGEEASVDSPWPQVSIQYDXQHVCVGSILDPHHVLTA	238
Qy	242	AHCFRKHDTDFNVKWVRAGSDKLGSFPFLAVAKIIIEFNPMYPKDNDIALMKLOPFLTFS	301
Dd	239	AHCFRKHDTDFNVKWVRAGSDKLGSFPFLAVAKIIIEFNPMYPKDNDIALMKLOPFLTFS	298
Qy	302	GTVRPICLPFFDEELTPATPLMIIGWGFTTKONGKMSDILLQASQVVIDSTRCNADDAVQ	361
Dd	299	GTVRPICLPFFDEELTPATPLMIIGWGFTTKONGKMSDILLQASQVVIDSTRCNADDAVQ	358
Qy	362	GEVTERRMCAGIPPEGGVDTCCGDSGGGPLMTQSQDWHVVGIVSWGYGCGGPFSTPGVVTKVS	421
Dd	359	GEVTERRMCAGIPPEGGVDTCCGDSGGGPLMTQSQDWHVVGIVSWGYGCGGPFSTPGVVTKVS	418
Qy	422	AYLNWIYNWKRAEL	435
Dd	419	AYLNWIYNWKRAEL	432

RESULT 12
 US-09-346-374-275
 ; Sequence 275, Application US/09946374
 ; Publication No. US20030073129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Botstein, David
 ; APPLICANT: Deenoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth J.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; TITLE OF INVENTION: Acids Encoding the Same
 ; FILE REFERENCE: F2830P1C1
 ; CURRENT APPLICATION NUMBER: US/09/946,374
 ; CURRENT FILING DATE: 2001-09-04
 ; PRIOR APPLICATION NUMBER: 60/098716
 ; PRIOR FILING DATE: 1998-09-01
 ; PRIOR APPLICATION NUMBER: 60/098723
 ; PRIOR FILING DATE: 1998-09-01
 ; PRIOR APPLICATION NUMBER: 60/098749
 ; PRIOR FILING DATE: 1998-09-01
 ; PRIOR APPLICATION NUMBER: 60/098750
 ; PRIOR FILING DATE: 1998-09-01
 ; PRIOR APPLICATION NUMBER: 60/098803
 ; PRIOR FILING DATE: 1998-09-02
 ; PRIOR APPLICATION NUMBER: 60/098821
 ; PRIOR FILING DATE: 1998-09-02

1	PRIOR FILING DATE: 1998-09-23	
2	PRIOR APPLICATION NUMBER: 60/101475	
3	PRIOR FILING DATE: 1998-09-23	
4	PRIOR APPLICATION NUMBER: 60/101476	
5	PRIOR FILING DATE: 1998-09-23	
6	PRIOR APPLICATION NUMBER: 60/101477	
7	PRIOR FILING DATE: 1998-09-23	
8	PRIOR APPLICATION NUMBER: 60/101479	
9	PRIOR FILING DATE: 1998-09-23	
10	PRIOR APPLICATION NUMBER: 60/101738	
11	PRIOR FILING DATE: 1998-09-24	
12	PRIOR APPLICATION NUMBER: 60/101741	
13	PRIOR FILING DATE: 1998-09-24	
14	PRIOR APPLICATION NUMBER: 60/101743	
15	PRIOR FILING DATE: 1998-09-24	
16	PRIOR APPLICATION NUMBER: 60/101915	
17	PRIOR FILING DATE: 1998-09-24	
18	PRIOR APPLICATION NUMBER: 60/101916	
19	PRIOR FILING DATE: 1998-09-24	
20	PRIOR APPLICATION NUMBER: 60/102207	
21	PRIOR FILING DATE: 1998-09-29	
22	PRIOR APPLICATION NUMBER: 60/102240	
23	PRIOR FILING DATE: 1998-09-29	
24	PRIOR APPLICATION NUMBER: 60/102307	
25	PRIOR FILING DATE: 1998-09-29	
26	PRIOR APPLICATION NUMBER: 60/102330	
27	PRIOR FILING DATE: 1998-09-29	
28	PRIOR APPLICATION NUMBER: 60/102331	
29	PRIOR FILING DATE: 1998-09-29	
30	PRIOR APPLICATION NUMBER: 60/102484	
31	PRIOR FILING DATE: 1998-09-30	
32	PRIOR APPLICATION NUMBER: 60/102487	
33	PRIOR FILING DATE: 1998-09-30	
34	PRIOR APPLICATION NUMBER: 60/102570	
35	PRIOR FILING DATE: 1998-09-30	
36	PRIOR APPLICATION NUMBER: 60/102571	
37	PRIOR FILING DATE: 1998-09-30	
38	PRIOR APPLICATION NUMBER: 60/102684	
39	PRIOR FILING DATE: 1998-10-01	
40	PRIOR APPLICATION NUMBER: 60/102687	
41	PRIOR FILING DATE: 1998-10-01	
42	PRIOR APPLICATION NUMBER: 60/102965	
43	PRIOR FILING DATE: 1998-10-02	
44	PRIOR APPLICATION NUMBER: 60/103258	
45	PRIOR FILING DATE: 1998-10-06	
46	PRIOR APPLICATION NUMBER: 60/103314	
47	PRIOR FILING DATE: 1998-10-07	
48	PRIOR APPLICATION NUMBER: 60/103315	
49	PRIOR FILING DATE: 1998-10-07	
50	PRIOR APPLICATION NUMBER: 60/103328	
51	PRIOR FILING DATE: 1998-10-07	
52	PRIOR APPLICATION NUMBER: 60/103395	
53	PRIOR FILING DATE: 1998-10-07	
54	PRIOR APPLICATION NUMBER: 60/103396	
55	PRIOR FILING DATE: 1998-10-07	
56	PRIOR APPLICATION NUMBER: 60/103401	
57	PRIOR FILING DATE: 1998-10-07	
58	PRIOR APPLICATION NUMBER: 60/103449	
59	PRIOR FILING DATE: 1998-10-06	
60	PRIOR APPLICATION NUMBER: 60/103633	
61	PRIOR FILING DATE: 1998-10-08	
62	PRIOR APPLICATION NUMBER: 60/103678	
63	PRIOR FILING DATE: 1998-10-08	
64	PRIOR APPLICATION NUMBER: 60/103679	
65	PRIOR FILING DATE: 1998-10-08	
66	PRIOR APPLICATION NUMBER: 60/103711	
67	PRIOR FILING DATE: 1998-10-08	
68	PRIOR APPLICATION NUMBER: 60/104257	
69	PRIOR FILING DATE: 1998-10-14	
70	PRIOR APPLICATION NUMBER: 60/104987	
71	PRIOR FILING DATE: 1998-10-20	
72	PRIOR APPLICATION NUMBER: 60/105000	
73	PRIOR FILING DATE: 1998-10-20	

; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 98.1%; Score 2297.5; DB 10; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222; Indels 5; Gaps 1;
Matches 429; Conservative 0; Mismatches 0

QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63

QY 62 CGQPLHFIPIKQDCGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPIKQDCGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 123

QY 122 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELRMENSSGPC 181
DB 124 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELRMENSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTA 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTA 238

QY 242 AHCERKHTDVFNWVKRAGSKLGSPPSLAVAKIIIEFNPMYKPDNDIALMKLOFPLTFS 301
DB 239 AHCERKHTDVFNWVKRAGSKLGSPPSLAVAKIIIEFNPMYKPDNDIALMKLOFPLTFS 298

QY 302 GTVRPCLPFDEELTPATPLWIIIGWFTKONGKGVSDILLQASVQVLDSTRCNADDAAYQ 361
DB 299 GTVRPCLPFDEELTPATPLWIIIGWFTKONGKGVSDILLQASVQVLDSTRCNADDAAYQ 358

QY 362 GEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPTGVYTKVS 421
DB 359 GEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPTGVYTKVS 418

QY 422 AYLNWYNNWKAEL 435
DB 419 AYLNWYNNWKAEL 432

RESULT 13
US-10-206-915-330
; Sequence 330, Application US/10206915
; Publication No. US20040029221A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1CS3
; CURRENT APPLICATION NUMBER: US/10/206,915
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586

; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-206-915-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222; Indels 5; Gaps 1;
Matches 429; Conservative 0; Mismatches 0

QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63

QY 62 CGQPLHFIPIKQDCGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPIKQDCGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 123

QY 122 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELRMENSSGPC 181
DB 124 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELRMENSSGPC 178

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTA 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLTA 238

QY 242 AHCERKHTDVFNWVKRAGSKLGSPPSLAVAKIIIEFNPMYKPDNDIALMKLOFPLTFS 301
DB 239 AHCERKHTDVFNWVKRAGSKLGSPPSLAVAKIIIEFNPMYKPDNDIALMKLOFPLTFS 298

QY 302 GTVRPCLPFDEELTPATPLWIIIGWFTKONGKGVSDILLQASVQVLDSTRCNADDAAYQ 361
DB 299 GTVRPCLPFDEELTPATPLWIIIGWFTKONGKGVSDILLQASVQVLDSTRCNADDAAYQ 358

QY 362 GEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPTGVYTKVS 421
DB 359 GEVTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPTGVYTKVS 418

QY 422 AYLNWYNNWKAEL 435
DB 419 AYLNWYNNWKAEL 432

RESULT 14
US-10-199-670-330
; Sequence 330, Application US/10199670
; Publication No. US20040033560A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C401
; CURRENT APPLICATION NUMBER: US/10/199,670
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-199-670-330

Query Match      98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY      2 DPDSOPLNSLDVKPRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 61
DB      4 DPDSOPLNSLDVKPRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 63

QY      62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSATGNW 121
DB      64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSATGNW 123

QY      122 FSACFDNFTALAEACROMGYSSKPTFRAVEIGDPDQDLVDVEITENSQELRMNSSGPC 181
DB      124 FSACFDNFTALAEACROMGYSSKPTFRAVEIGDPDQDLVDVEITENSQELRMNSSGPC 178

QY      182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
DB      179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 238

QY      242 AHCFRKHDTVFNWKRAGSDKLGSPSPSLAVAKIIIEFNPMYKDNIDALMKLOPPLTFS 301
DB      239 AHCFRKHDTVFNWKRAGSDKLGSPSPSLAVAKIIIEFNPMYKDNIDALMKLOPPLTFS 298

QY      302 GTVRICLPFFDEELTPATPLMIIGWGTCKONGKMSDILLQASVQVLDSTRCNADDAQY 361
DB      299 GTVRICLPFFDEELTPATPLMIIGWGTCKONGKMSDILLQASVQVLDSTRCNADDAQY 358

QY      362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGCGGPGTGPVYTKVS 421
DB      359 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGCGGPGTGPVYTKVS 418

QY      422 AYLNWIYNVWKAE 435
```

```

DB      419 AYLNWIYNVWKAE 432

RESULT 15
US-10-201-858-330
; Sequence 330, Application US/10201858
; Publication No. US20040038337A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C464
; CURRENT APPLICATION NUMBER: US/10/201,858
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-201-858-330

Query Match      98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY      2 DPDSOPLNSLDVKPRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 61
DB      4 DPDSOPLNSLDVKPRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 63

QY      62 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSATGNW 121
DB      64 CGQPLHFIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAPAVRLSKDRSTLQVLDSATGNW 123

QY      122 FSACFDNFTALAEACROMGYSSKPTFRAVEIGDPDQDLVDVEITENSQELRMNSSGPC 181
DB      124 FSACFDNFTALAEACROMGYSSKPTFRAVEIGDPDQDLVDVEITENSQELRMNSSGPC 178

QY      182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
DB      179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 238
```

QY	242	AHCFRKHTDVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS	301
Db	239	AHCFRKHTDVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS	298
QY	302	GTVRPICLFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ	361
Db	299	GTVRPICLFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ	358
QY	362	GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGIVSWGCGGSPSTPGVYTKVS	421
Db	359	GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVVGIVSWGCGGSPSTPGVYTKVS	418
QY	422	AYLNWIYNVWKAEL	435
Db	419	AYLNWIYNVWKAEL	432

Search completed: June 7, 2004, 08:47:29
Job time : 49 secs

This Page Blank (uspto)